



**Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

Dirección General de Estudios de Posgrado

Facultad de Ciencias Biológicas

Unidad de Posgrado

**Análisis genómico de *Mycobacterium tuberculosis*  
sensible, MDR y XDR aislados en el Perú**

**TESIS**

Para optar el Grado Académico de Magíster en Biología  
Molecular

**AUTOR**

David Demetrio TARAZONA JURADO

**ASESOR**

Dr. Abelardo Lenin MATURRANO HERNANDEZ

Lima, Perú

2019



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Tarazona, D. (2019). *Análisis genómico de Mycobacterium tuberculosis sensible, MDR y XDR aislados en el Perú*. Tesis para optar grado de Magíster en Biología Molecular. Unidad de Posgrado, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

---



**Universidad Nacional Mayor de San Marcos**  
Universidad del Perú. Decana de América

Vicerrectorado de Investigación y Posgrado  
Dirección General de Biblioteca y Publicaciones

Dirección del Sistema de Bibliotecas y Biblioteca Central



"Año de la lucha contra la corrupción y la impunidad"

## **Hoja de metadatos complementarios**

Código ORCID del autor: David Demetrio Tarazona Jurado

0000-0002-9273-0590

Código ORCID del asesor: Abelardo Lenin Maturrano Hernández

0000-0001-8819-7335

DNI del autor: 42681920

Grupo de investigación:

Institución que financia parcial o totalmente la investigación:

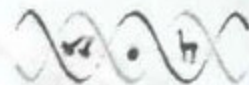
Instituto Nacional de Salud

Ubicación geográfica donde se desarrolló la investigación. Debe incluir localidades y/o coordenadas geográficas:

Lima, Lima

Año o rango de años que la investigación abarcó:

2016



FACULTAD DE CIENCIAS BIOLÓGICAS

VICEDECANATO DE INVESTIGACION Y POSGRADO

UNIDAD DE POSGRADO

Exped. N° 158-UPG-FCB-2017

ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR AL GRADO ACADÉMICO DE  
MAGÍSTER EN BIOLOGÍA MOLECULAR

Siendo las 14:10 horas del día 4/9/2019 en el Salón de Grados de la Facultad de Ciencias Biológicas, el Jurado de Tesis presidido por:

Dra. Susana Mónica Gutiérrez Moreno	e integrado por
Dra. Egma Marcelina Mayta Huatuco	(Miembro)
Mg. Ruth Hortensia García de la Guarda	(Miembro)
Dr. Dan Erick Vivas Ruíz	(Miembro)
Dr. Abelardo Lenin Maturrano Hernández	(Asesor)

Se reunió para la sustentación oral y pública de la Tesis para optar al Grado Académico de Magíster en Biología Molecular, que solicitara el Bachiller Don **DAVID DEMETRIO TARAZONA JURADO**.

Después de darse lectura al Expediente N° 158-UPG-FCB-17, en el que consta haberse cumplido con todas las disposiciones reglamentarias, los señores miembros del Jurado, recepcionaron la exposición de la Tesis Titulada:

**"ANÁLISIS GENÓMICO DE *Mycobacterium tuberculosis* SENSIBLE, MDR y XDR AISLADOS EN EL PERÚ"**, formuladas las preguntas, éstas fueron absueltas por el graduando.



VICEDECANATO DE INVESTIGACION Y POSGRADO

UNIDAD DE POSGRADO

Acto seguido el Jurado procedió a la votación la que dio como resultado el calificativo de: *Muy Bueno con la nota de 14*

A continuación el Presidente del Jurado de Tesis recomienda que la Facultad proponga que la Universidad le otorgue el Grado Académico de Magíster en Biología Molecular al señor Bachiller **DAVID DEMETRIO TARAZONA JURADO**.

Siendo las *19:30* se levantó la Sesión, recibiendo el graduando las felicitaciones de los señores miembros del Jurado y público asistente.

Se extiende la presente Acta en Lima, a los 04 días del mes de setiembre del año 2019.

**Dra. Susana M. Gutiérrez Moreno**  
Profesora Principal a T.C.  
PRESIDENTE  
REGINA N° 13548

**Dr. Abelardo L. Maturrano Hernández**  
Profesor Principal a D.E.  
ASESOR  
REGINA N° 14184

**Dra. Egma M. Mayta Huatucu**  
Profesora Principal a D. E.  
MIEMBRO  
REGINA N° 13619

**Mg. Ruth H. García de la Guard**  
Profesora Principal a D. E.  
MIEMBRO  
REGINA N° 15295

**Dr. Dán-Erick Vivas Ruiz**  
Profesor Asociado a T. C.  
MIEMBRO  
REGINA N° 13418

Agradezco a Dios por todas las cosas,

A mi familia, por brindarme apoyo constante

A mi mentor, Dr. Lenin, quien me inicio en la genómica

A mi mentor, Dr. Heinner, por sus sabios consejos en la academia

Y a mis profesores de UNMSM por

transformar a la ciencia en un amigo de toda la vida.

## CONTENIDO

Resumen	1
Abstract	1
I. INTRODUCCIÓN	1
II. ANTECEDENTES	6
1. La tuberculosis en la antigüedad	6
2. La aparición de <i>Mycobacterium tuberculosis</i> como patógeno humano	7
3. La epidemiología de la tuberculosis	8
4. Genotipos de <i>Mycobacterium tuberculosis</i>	10
5. La drogoresistencia en <i>Mycobacterium tuberculosis</i>	12
6. La genómica de <i>Mycobacterium tuberculosis</i> para el diagnóstico	33
III. HIPÓTESIS Y OBJETIVOS	35
IV. MATERIAL Y MÉTODOS	36
1. Cepas, sensibilidad a drogas y purificación de ADN	36
2. Datos genómicos de cepas de <i>Mycobacterium tuberculosis</i>	36
3. Linaje y secuenciamiento genómico	37
4. Ensamblaje y anotación	37
5. Identificación de SNP, COG y análisis comparativo	38
V. RESULTADOS	40
1. Estructura de los genomas de aislados peruanos de <i>M. tuberculosis</i>	40
2. Comparación de secuencias de aislados peruanos <i>M. tuberculosis</i>	42
3. Análisis de grupos ortólogos	43
VI. DISCUSIÓN	47
VII. CONCLUSIONES	51
VIII. RECOMENDACIONES	52
IX. REFERENCIAS BIBLIOGRÁFICAS	53
X. Anexo 1 - Functional Analysis Using Whole-Genome Sequencing of a Drug-Sensitive <i>Mycobacterium tuberculosis</i> Strain from Peru	64



XI.	Anexo 2 - Evidence of Clonal Expansion in the Genome of a Multidrug-Resistant <i>Mycobacterium tuberculosis</i> Clinical Isolate from Peru	67
XII.	Anexo 3 - Genome Analysis of 17 Extensively Drug-Resistant Strains Reveals New Potential Mutations for Resistance	70
XIII.	Anexo 4 - Análisis genómico comparativo de cepas peruanas de <i>Mycobacterium tuberculosis</i>	73
XIV.	Anexo 5. Variantes genéticas de <i>Mycobacterium tuberculosis</i> INS_SEN1	86
XV.	Anexo 6. Variantes genéticas de <i>Mycobacterium tuberculosis</i> INS_MDR2	108
XVI.	Anexo 7. Variantes genéticas de <i>Mycobacterium tuberculosis</i> INS_XDR3	138

## RESUMEN

La tuberculosis (TB) es una de las enfermedades más importantes en Salud Pública en el mundo y en el Perú tiene una notable importancia por el número de casos de TB registrados y el incremento de casos drogorresistentes en los últimos años. Se realizó el análisis genómico comparativo de tres genomas de *Mycobacterium tuberculosis* que mostraban características fenotípicas de sensibilidad a drogas (INS-SEN), multidrogorresistente (INS-MDR) y extremadamente resistente (INS-XDR). Se determinó que los aislados de INS-MDR y de INS-XDR difieren en 6.1% SNPs, mientras que el análisis comparativo del genoma del aislado INS-SEN con la cepa INS-MDR y la cepa INS-SEN con INS-XDR difieren en 50.2% y 50.3% respectivamente. Las variaciones en el genoma de *M. tuberculosis* en los aislados INS-SEN, INS-MDR y INS-XDR con más frecuencia pertenecen a la familia de genes Q (biosíntesis, catabolismo y transporte de metabolitos secundarios) y genes L (recombinación, replicación y reparación del ADN). Se concluye que el aislado sensible y drogorresistentes que incluye MDR y XDR no se encuentran relacionados estrechamente.

Palabras clave. Tuberculosis, genoma y drogoresistencia.

## **ABSTRACT**

Tuberculosis (TB) is one of the most important diseases in Public Health in the world and Peru has a remarkable importance due to the number of registered TB cases and the increase of drug resistant cases in recent years. We performed a comparative genomic analysis of three *Mycobacterium tuberculosis* genomes that showed phenotypic characteristics of drug sensitivity (INS-SEN), multi-drug resistant (INS-MDR) and extremely resistant (INS-XDR). It was determined that the INS-MDR and INS-XDR isolates differ in 6.1% SNPs, while the comparative analysis of the INS-SEN isolate genome with INS-MDR and INS-SEN with INS-XDR differ by 50.2% and 50.3% respectively. Variations in the genome of *M. tuberculosis* in INS-SEN, INS-MDR and INS-XDR isolates are most frequently in the Q gene family (biosynthesis, transport and secondary metabolism catabolism) and L genes (replication, recombination and repair). There is evidence that in the case of drug-sensitive and drug-resistant isolates (MDR and XDR) they are not closely related.

Key words. Tuberculosis, Genome, Drug Resistant

## I. INTRODUCCIÓN:

La tuberculosis (TB) es causada por *Mycobacterium tuberculosis* (*Mtb*), una enfermedad que casi siempre afecta a los pulmones en los seres humanos. Durante el año 2013, según la OMS, se tiene alcanzado 6.1 millones de nuevos casos de pacientes con tuberculosis. Un 3.5% corresponden a nuevos pacientes con tuberculosis multidrogorresistente (TB-MDR) y un 9% han formado nuevos casos de tuberculosis extremadamente drogorresistente (TB-XDR). En el Perú, una elevada carga de pacientes con tuberculosis y el aumento de casos con multidrogorresistentes (MDR) constituye uno de los principales problemas en salud pública. Desde el 2013 la incidencia de pacientes con tuberculosis en el Perú alcanzo 124/100,000 habitantes, de ello un 3.9% (n=850) existieron nuevos pacientes con diagnostico TB-MDR (WHO., 2014).

Los nuevos casos de TB se han reducido recientemente, sin embargo, el numero de pacientes con diagnostico TB-MDR van en aumento desde los últimos 5 años. En el año 2014, Perú constituyó un 41.3% del total de los pacientes con TB-MDR (Murray et al., 2014), siendo la ciudad de Lima, la que ha llegado a concentrar alrededor del 85% y 92% de reportes de pacientes con TB-MDR y TB-XDR, respectivamente (O. Caceres et al., 2014). En Lima, se tiene evidencia que el principal elemento de riesgo para pacientes con TB y su propagación es debido al transporte público peruano (Garaycochea & Ticona, 2015; Horna-Campos, Sanchez-Perez, Sanchez, Bedoya, & Martin, 2007), lo que podría contribuir a la continuidad del aumento de casos de TB permanentemente.

La definición de TB-MDR esta determinada como la TB resistente a los antibioticos como rifampicina (RIF) e isoniacida (INH), mientras TB-XDR está definida como la TB resistente a rifampicina, isoniazida, una fluoroquinolona (FLQ) y un antibiótico de segunda línea (capreomicina, kanamicina, amikacina).

La Rifampicina es un tipo de antibiótico con actividad bactericida que tiene la capacidad de inhibir la enzima ARN polimerasa de bacterias. La región del ADN asociada a resistencia de RIF está identificada en un fragmento de 81 pb correspondiente al gen *rpoB* (Subunidad Beta de ARN polimerasa de *Mtb*). La Isoniacida es una pro-droga que se activa mediante la enzima *katG* (catalasa peroxidasa) del microbio *Mtb* por inhibición de la síntesis del ácido micólico. La resistencia a la droga de INH es frecuentemente determinada en el codón 315 de la secuencia del gen *katG*, y otras significativas mutaciones de los genes *kasA*, *ndh*, *ahpC* y la sección del promotor de *inhA* (Akos Somoskovi, Linda M Parsons, & Max Salfinger, 2001). Las fluoroquinolonas son componentes antimicrobianos que tienen un rol en la inhibición de ADN girasa (topoisomerasa II) de *Mtb*, codificadas por los genes *gyrA* y *gyrB* que comprende de dos subunidades A y B. El mecanismo genético para la resistencia a FLQ es por los cambios en el ADN girasa, específicamente mutaciones de *gyrA* (Soudani et al., 2010), las más comunes encontradas se encuentran en la Region Determinante de Resistencia a Quinolona (QRDR) en el *gyrA* (Pantel et al., 2011), las mutaciones en *gyrB* pueden tambien asociarse con la resistencia a la FQ, pero con reducida sensibilidad y especificidad, y que frecuentemente ocurre paralelamente con mutaciones en *gyrA* (Z. Cui, Wang, Lu, Huang, & Hu, 2011; Duong et al., 2009). Aunque la mayoría de las cepas *Mtb* con mutaciones QRDR en *gyrA* son resistentes a FLQ, casi todos los aislados con un QRDR tipo salvaje son susceptibles FLQ.

La gran diversidad genómica de *Mtb* a nivel global son las responsables de causar la enfermedad de tuberculosis. A nivel mundial las cepas más prevalentes incluyen a los genotipos: Latino América-Mediterráneo (LAM), Haarlem (H), Este de África-India (EAI), T, Beijing, Central-Asia (CAS) y X. La asignación de los genotipos está asociado al lugar de origen donde fue aislada, identificada y caracterizada inicialmente (Philip Supply et al., 2006). Así mismo, las recientes investigaciones en análisis genómicos completos están ayudando a identificar regiones de *Mtb* para potenciales posiciones polimórficas en asociación a resistencia a drogas antituberculosas (Casali et al., 2012; Ford et al., 2013; Timothy M Walker et al., 2013).

Las cepas MDR y XDR tienen orígenes principalmente desde el sur de África, las cuales incluyen a las principales familias F11 y KZN. Las cepas del grupo F11 fueron detectadas en los años 1993-1999, en Cape - África y constituyó un 21.4% de casos de TB reportados ese año. Sus tipologías incluyen la presencia de la región *IS6110* la cual modifica de 11 a 19 bandas, deleción de regiones espaciadoras 9-11, 21- 24 y 33-36 y finalmente la mutación en C491T del gen *rrs* (Victor et al., 2004). El grupo de KZN procede de KwaZulu-Natal del país de Sudáfrica derivado de un brote mortal entre 2005-2007. Es un tipo de cepa con alta virulencia, alcanzó 52 muertes de los 53 casos con TB-XDR y HIV. Estas cepas fueron secuenciadas genéticamente en el año 2007 y están disponibles en la base de datos Genbank, pero presentan regiones incompletas (Ioerger et al., 2009).

En el Perú, las poblaciones con ascendencia nativa han sido infectados por TB (Culqui et al., 2010), pero ese linaje de *Mtb* que los afectó aún permanece desconocido. Los registros peruanos indican que la familias Beijing empezaron desde la llegada de

migrantes asiáticos (Iwamoto et al., 2012), lo que podríamos decir como importación de cepas de *Mtb*. Las migraciones de Peruanos a otras partes del mundo, como a España y Italia ha ocasionado que los genotipos de *Mtb* pueda haber sido transferida a países Europeos (García de Viedma, Chaves, & Inigo, 2006), un tipo de exportación de cepas de *Mtb*.

Uno de los primeros reportes en el Perú sobre diversidad genómica de *Mtb* de aislados sensible y drogoresistentes identificó los genotipos frecuentes: T (23.8%), LAM (23.8%), Haarlem (22.3%) y Beijing (9.3%) (Taype et al., 2012). Sheen *et al.* reportaron en pacientes coinfectados con TB y VIH positivos como prevalentes a los linajes Haarlem (28.7%), LAM (28.3%) y T (20.3%) (Sheen et al., 2013). Cáceres *et al.*, reportaron *Mtb* de genotipos Haarlem (43.6%), T (27.7%), LAM (16.2%) y Beijing (9.5%) en pacientes con TB-XDR (O. Cáceres et al., 2014). Estas investigaciones permitieron un primer intento de la diversidad de genotipos LAM, Haarlem y T presentes en Perú. Sin embargo, más recientemente, Grandjean *et al.*, en una investigación con 2,139 aislados de *Mtb* en la región peruana, determinó que el genotipo familia LAM puede constituir como mínimo el 50% de los casos reportados de pacientes con resistencia a droga (Grandjean et al., 2015), esto puede deberse a que LAM es la más antigua y con mayor prevalencia a nivel global.

Durante el año 2015 en Perú, se han publicado genomas completos de *Mtb* sensible, MDR XDR y en el 2016 se realizó un análisis comparativo entre los genomas, sin embargo se requiere más datos de secuenciación y análisis de genomas. La

presente tesis tiene como objetivo analizar comparativamente 3 genomas de *Mtb* entre ellos y frente a otros genomas referenciales.



## **II. ANTECEDENTES**

### **1. La tuberculosis en la antigüedad**

Solo algunas enfermedades han tenido una asociación con los seres humanos que es tan antigua e impactante como la tuberculosis. Descripciones literarias que son sugestivos a tuberculosis son escritos médicos de China (~ 2700 aC), India (~ 1500 aC), en la Biblia en el Antiguo Testamento (~1,300-400 aC) y los escritos de Hipócrates (460 aC) los cuales también refieren a distintos nombres que describen la sintomatología de tuberculosis pulmonar (Galagan, 2014). La asignación del nombre tuberculosis es debido a una asociación fisiopatológico para el desarrollo de un granuloma o tubérculo. Esta antigua asociación de la tuberculosis y la humanidad ha sido apoyada por técnicas moleculares y datos arqueológicos. Aunque las infecciones de tuberculosis son más comunes en los pulmones (85%), la tuberculosis podría llegar a presentarse en prácticamente cualquier parte del cuerpo (la cual está limitada por el uso actual de la vacuna).

Las investigaciones arqueológicas de la tuberculosis se han aplicado ampliamente a las antiguas momias egipcias (Crubezy et al., 1998), por lo que se ha podido estimar que el 25% de los egipcios pudieron haber sufrido de tuberculosis (Nerlich & Losch, 2009). Incluso ha sido reportado desde el primer milenio en China y otros países como Alemania, Italia, Suecia y Dinamarca (Galagan, 2014).

La tuberculosis humana tiene como evidencia fósil más antigua correspond hasta los 9.000 de edad en el Mediterráneo (Hershkovitz et al., 2008), esto sugiere una versión

moderna de la tuberculosis durante el período neolítico. Esto coloca en evidencia que la tuberculosis humana es sincrónico con la domesticación de animales y considera a lugar la hipótesis que la infección por *Mtb* surgió a través de una transmisión zoonótica de una cepa de *M. bovis* de bovino (Bates & Stead, 1993; Kapur, Whittam, & Musser, 1994). La identificación rastros moleculares de *Mtb* en restos humanos precolombinos Peruanos (Salo, Aufderheide, Buikstra, & Holcomb, 1994) y Chilenos (Arriaza, Salo, Aufderheide, & Holcomb, 1995) sugieres que miembros portadores de *Mtb* pudieron haber cruzado al nuevo mundo a través del estrecho de Bering en el Pleistoceno, antes de la domesticación de animales. Por otra parte, las lesiones típicas de la tuberculosis se han identificado en América del Norte (hace 125,000-8,000 años) y mastodontes (hace 38,000-10,000 años) reveló que el ADN de micobacterias está más estrechamente alineada con una moderna *Mtb* que a *M. africanum* y *M. bovis*. Estos datos arqueológicos sugieren que la tuberculosis brotó como una zoonosis, posiblemente anterior a la domesticación de animales (Galagan, 2014). Por lo tanto, la Tuberculosis nos ha acompañado desde hace miles de años, pero nuestra lucha directa contra la TB recién ha comenzado 50 años atrás.

## **2. La aparición de *Mycobacterium tuberculosis* como patógeno humano**

Durante miles de años la tuberculosis ha afectado a los seres humanos. La humanidad y la Tuberculosis ha sido integrada por evidencia morfológica en fósil humano, documentación literaria y la identificación de ADN micobacterial en restos humanos usando PCR (Crubezy, Legal, Fabas, Dabernat, & Ludes, 2006; Zink, Grabner, & Nerlich, 2005).

Los análisis iniciales indican que la infección humana de *Mtb* surgió a través de una transmisión zoonótica de una antigua cepa de *M. bovis* de vacas durante la domesticación (Labidi & Thoen, 1989), similar a los orígenes zoonóticos de otros patógenos como el sarampión y la gripe. Sin embargo, los estudios genómicos de otros miembros de *Mtb* y bacilos tuberculosos han permitido reformular estas hipótesis. Utilizando el análisis de reloj molecular, se sugiere que el ancestro de *Mtb* y bacilos tuberculosos de forma lisa (BTL), podrían haber existido hace 2.8 millones de años (Gutierrez et al., 2005). Lo que supone que la enfermedad por micobacterias puede ser incluso anterior al *Homo sapiens* lo que podría haber infectado a antepasados humanos. Las investigaciones de BTL indican que la aparición de la patogenicidad se relacionó con un cambio en la evolución debido a que las cepas BTL son genéticamente variables lo que evidencia la recombinación y la existencia de transferencia lateral de genes (Mortimer & Pepperell, 2014). Por el contrario, las *Mtb* parece ser la única expansión clonal de esta población de progenitores. En general los agentes de *Mtb* tienen una reducida diversidad genética y no se ha reportado evidencia de transferencia horizontal de genes.

### **3. La epidemiología de la tuberculosis**

En 2015 a nivel mundial se estimó 10,4 millones de nuevos casos (incidentes) de TB, de los cuales 1,0 millón (10%) niños, 3,4 millones (33%) fueron mujeres y 5,8 millones (54%) fueron hombres (Dean, Cox, & Zignol, 2017). El 61 % de nuevos reportes se dieron en seis países: China, Indonesia, India, Pakistán, Sudáfrica y Nigeria. La tasa de reducción de la incidencia mundial de la TB de 2014 a 2015 se conservó en tan solo un 1,5% (Dean et al., 2017). En el 2015 los nuevos casos de TB-MDR fue 480'000

y más de 100'000 reportes de TB resistentes a las drogas antituberculosas que incluyen rifampicina, para el tratamiento de la TB-MDR. La Federación de Rusia, la India y China representaron el 45% de 580'000 casos. En el 2015 los casos de muertes por TB alcanzo los 1,51 millones y adicionalmente 0,39 millones de muertes por TB en personas infectocantagiadas por VIH. Las muertes por TB se han reducido en un 22% entre los años 2000 y 2015 (Dean et al., 2017).

La TB en el Perú equivale al décimo quinto lugar de causas de muerte, y el vigésimo séptimo puesto de carga de tuberculosis medidos por años de vida saludable que han sido perdidos. Esta enfermedad afecta, principalmente a los estratos sociales más pobres de las grandes ciudades del Peru. Las tasas informadas de incidencia (casos nunca tratados por cada 100 mil habitantes) y morbilidad (nuevos y tratados por cada 100 mil habitantes) se han reducido hasta 3% por año entre 2011 al 2015, de 109,7 a 99,5 en morbilidad y 97,4 a 87,6 en incidencia (Valentina Alarcón, 2017).

En el año 2015 se informaron 30 988 casos de TB, y la tasa de incidencia fue la menor reportada desde hace 25 años, ascendiendo solo a 87,6 casos nuevos de TB por cada 100 mil habitantes. La ESNPCT notifica las personas fallecidas durante el tratamiento, y se calcula como sustituto de la tasa de mortalidad por cada 100 mil habitantes, la que se mantiene en 3,7 de los últimos 3 años (Valentina Alarcón, 2017).

#### **4. Genotipos de *Mycobacterium tuberculosis***

Los dos principales métodos de genotipificación son basados en repeticiones palíndromas cortas agrupadas y regularmente espaciadas (CRISPR del inglés, Clustered Regulatory Short Palindromic Repeats) y número variable de repeticiones en tándem (VNTR del inglés Variable Number Tandem Repeats). Los métodos de genotipificación son conocidos como espoligotipos basados en identificación de CRISPR y 'MIRU-VNTR' basados en VNTR. El método espoligotipos es una técnica basada en la detección de 43 posiciones de espaciadores únicos repetidos en un locus específico, también conocido como repetición directa (RD). Los patrones de espoligotipos son generalmente representados por bloques negros y blancos indicando presencia o ausencia de un espaciador en particular (Kato-Maeda, Metcalfe, & Flores, 2011).

En *Mtb* se han identificado 14 regiones diferenciales (RD1-14), que contienen un rango de tamaño de 2- 12.7kb, que se han encontrado ausentes en Bacilo Calmette-Guerin Pasteur, relativo a *Mtb* H37Rv(Behr et al., 1999; Gordon et al., 1999). Sin embargo, en diversos estudios determinaron que RD3 y RD11 corresponden a dos profagos phiRv1 y phiRv2 de *Mtb* H37Rv(Cole et al., 1998), la región RD6 tiene inserción IS1532, y la región RD5 es flanqueado por una copia de IS6110 (Gordon et al., 1999), las regiones RvD2-RvD3, RvD4 y RvD5 presentan una gran variabilidad, debido a que algunas cepas pueden tener un alto número de copias de IS6110 (>14) por lo que estas regiones RvD pueden perderse, sin embargo estas diferencias no se correlacionan con la droga resistencia en *M. tuberculosis*.

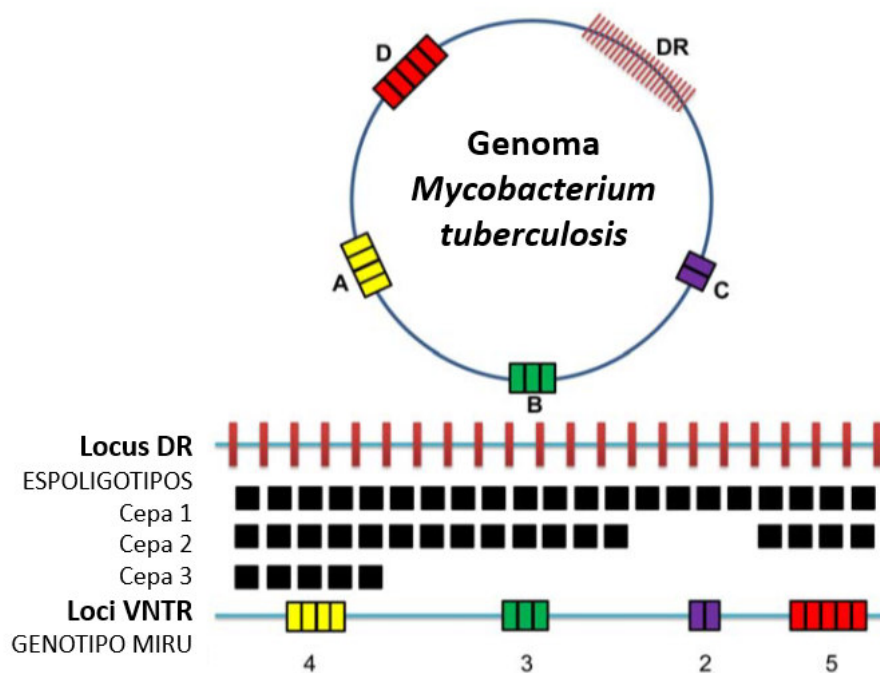


Figura 1. Esquema de genotipificación basado en CRISPR y VNTR en *M. tuberculosis*. Adaptación (Comas, Homolka, Niemann, & Gagneux, 2009).

Es posible categorizar a *Mtb* como aislados antiguos o modernos, basados en la presencia o ausencia de una región TbD1. Las que pueden ser categorizadas en Haarlem, Beijing entre otras. Sin embargo un estudio encontró que la región TbD1 fue ausente, hasta un 87% de los aislados más representativos como Haarlem, Beijing y clúster de África (Kremer et al., 1999). Incluso TbD1 no es exclusivo de *Mtb*, pues recientemente con el secuenciamiento genómico *M. canetti*, *M. africanum* y *M. bovis* (Garnier et al., 2003) también se ha descubierto que otras micobacterias presentan esta región.

## 5. Drogoresistencia en *Mycobacterium tuberculosis*

La bacteria *Mycobacterium tuberculosis* es hoy en día considerado como uno de los patógenos más exitosos y infecta a un tercio de la población mundial (Corbett et al., 2003). La bacteria tiene una capacidad innata para sobrevivir de los mecanismos de defensa del huésped, además *Mtb* puede resistir diferentes agentes antimicrobianos comercialmente disponibles (Nguyen & Pieters, 2009). Lo que origina que las opciones quimioterapéuticas comerciales para el tratamiento de la tuberculosis puedan ser limitadas severamente (Tabla 1).

Tabla 1. Drogas antituberculosas

<b><i>Drogas de primera línea</i></b>	<b><i>Drogas de segunda línea</i></b>
Isoniacida	Cicloserina
Rifampicina	Etionamida
Rifabutina	Ácido aminosalicílico (PAS)
Rifapentina	Estreptomina
Pirazinamida	Amikacina
Etambutol	Kanamicina
-	Capreomicina

Los regímenes prolongados utilizando los mismos pocos fármacos han resultado en el cumplimiento deficiente en pacientes y como resultado, la aparición de cepas que son cada vez más resistentes a los fármacos anti-TB disponibles. El uso desproporcionado de antibióticos ha generado una presión selectiva que estimula una rápida evolución de *M. tuberculosis*, empezando desde monodrogo resistente hasta

los resistentes a múltiples fármacos, debido a una acumulación secuencial de mutaciones genéticas de resistencia (O. Caceres et al., 2014; Dorman & Chaisson, 2007; Ormerod, 2005; Udwadia, 2012), en el Perú no se ha registrado casos de *Mtb* TDR. Las infecciones con *Mtb* XDR o TDR son fundamentalmente incurables por los actuales fármacos antituberculosos disponibles en el mercado, esto puede originar un riesgo en la salud pública global. Por ello se requieren nuevas terapias farmacológicas para hacer frente a la actual epidemia de TB resistente. Además del desarrollo de nuevas clases de fármacos antituberculosos, los enfoques no tradicionales. El éxito de estos enfoques requeriría una mejor comprensión de los mecanismos moleculares que subyace resistencia a fármacos en *Mtb*.

La resistencia a los antibióticos adquirida puede ocurrir en cualquiera de las bacterias a través de mutaciones o transferencia horizontal de genes (THG) mediada por fagos, plásmidos o transposones. Sin embargo en *Mtb* la THG para la resistencia a fármacos aun no ha sido reportada; la resistencia es principalmente derivada por mutaciones genéticas bajo una presión selectiva por utilización de antibióticos. Los genes mutados de *Mtb* que confieren resistencia a las drogas antituberculosas se enumeran en la Tabla 2.



Tabla 2. Genes implicados en la resistencia a los medicamentos adquiridos en *M. tuberculosis*.

Gen	Proteína Codificada	Función De Las Proteínas	Droga	Modo De Acción De Las Drogas
<b><i>kat G</i></b>	catalasa-peroxidasa	Activación profármaco	Isoniazida	La inhibición de la biosíntesis de ácidos micólicos y otros procesos metabólicos
<b><i>inh A</i></b>	Enoil reductasa	Objetivo de drogas		
<b><i>NDH</i></b>	deshidrogenasa II NADH	La modulación de la relación NADH / NAD		
<b><i>ahp C</i></b>	Alquil hidropoxidasa	La resistencia al estrés oxidativo		
<b><i>rpo B</i></b>	$\beta$ -subunidad de la ARN polimerasa	Objetivo de drogas	Rifampicina	La inhibición de la transcripción
<b><i>pnc A</i></b>	pirazinamidas	La activación profármaco	Pirazinamida	La inhibición de trans-traducción
<b><i>RSP A</i></b>	proteína ribosomal S1	Objetivo de drogas		
<b><i>embC AB</i></b>	Arabinosil transferasa	Objetivo de drogas	Etambutol	La inhibición de la síntesis de arabinogalactano
<b><i>EM BR</i></b>	regulador de la transcripción <i>emb CAB</i>	Regulación de la expresión <i>embcab</i>		
<b><i>rpsL</i></b>	proteína ribosomal S12	Objetivo de drogas	Estreptomina	La inhibición de la síntesis de proteínas
<b><i>RR</i></b>	16S rRNA	Objetivo de drogas		
<b><i>gid B</i></b>	16S rRNA metiltransferasa	Modificación del objetivo de drogas		
<b><i>whi B7</i></b>	Regulador de la transcripción MDR	Regulación de los genes de resistencia a fármacos		
<b><i>rrs</i></b>	ARN ribosómico 16S	Objetivo de drogas	Amikacina / kanamicina	La inhibición de la síntesis de proteínas
<b><i>eis</i></b>	aminoglucósido acetiltransferasa	Inactivación de drogas		
<b><i>whi B7</i></b>	regulador de la transcripción MDR	Regulación de la expresión <i>eis</i>		
<b><i>eth A</i></b>	Flavin monooxigenasa	La activación profármaco	Etionamida	La inhibición de la biosíntesis de micolato
<b><i>eth R</i></b>	represor de la transcripción <i>ethA</i>	Regulación de la expresión <i>etha</i>		
<b><i>inh</i></b>	Enoil reductasa ACP	Objetivo de drogas		
<b><i>NDH</i></b>	NADH deshidrogenasa II	La modulación de la relación NADH / NAD		

<b>MS HA</b>	glicosiltransferasa	La activación profármaco		
<b>gyr A</b>	ADN girasa subunidad A	Objetivo de drogas	Fluoroqui nolonas	La inhibición de la ADN girasa
<b>gyr B</b>	ADN girasa subunidad B	Unión de drogas		
<b>cyc A</b>	alanina ligasa	Objetivo de drogas	Cycloserin a	La inhibición de la síntesis del peptidoglicano
<b>ddl</b>	Acido Amino transportador	La captación del fármaco		
<b>thy A</b>	Timidilato sintasa	La síntesis de dttp	Ácido aminosalic ílico p(PA S)	La inhibición de la biosíntesis de folato
<b>DFR A</b>	Dihidrofolato reductasa	Objetivo de drogas		
<b>Folc</b>	Dihidrofolato sintasa	La activación profármaco		
<b>ribD</b>	Análogo de la dihidrofolato reductasa	La sustitución de la actividad objetivo de drogas		
<b>tlyA</b>	Metiltransferasa rRNA	Protección ribosoma	Capreomi cina	La inhibición de la síntesis de proteínas
<b>rrs</b>	ARN ribosómico 16S	Objetivo de drogas		

La progresión de la resistencia a los medicamentos contra *Mtb* resume la teoría de la evolución de Darwin, en la que la resistencia a los antibióticos se convierte en rasgos predominantes en poblaciones de *Mtb* porque traen ventajas de supervivencia de los mutantes que surgen bajo presión selectiva. La exposición del fármaco prolongada durante los regímenes, asociado a la falta de adherencia en los pacientes con diagnóstico de tuberculosis, ha estimulado a la microevolución para seleccionar mutantes genéticos resistentes que de otra circunstancia no predominarían en la población por su reducida eficacia biológica. Este proceso producido por la combinación de diferentes terapias ha generado una evolución de las cepas de *Mtb*, más resistentes a los fármacos comercialmente disponibles. Los medicamentos para la tuberculosis como isoniazida y la etionamida son profármacos, o sea requieren la activación de enzimas en el citoplasma de las micobacterias para que puedan tener actividad citotóxica. Sin embargo, si estos reactivos del oxígeno y radicales no pueden

matar a la célula micobacteriana, favorecerán la aparición de mutaciones de resistencia a fármacos.

### 5.1 RIFAMPICINA

La rifampicina es un fármaco derivado de la rifamicina, también un agente antituberculoso. Está catalogado entre los antibióticos más eficaces frente a la tuberculosis que integrado con isoniazida constituyen la base para el tratamiento de pacientes con tuberculosis multirresistente. La rifampicina es activa en bacilos en crecimiento y en no crecimiento (metabolismo lento) (Mitchison, 1979). El mecanismo de acción de rifampicina en *Mtb* es por la unión de la subunidad  $\beta$  de la ARN polimerasa lo que origina la inhibición de la elongación de la molecular del ARN mensajero (Blanchard, 1996). Generalmente, los aislados clínicos con resistencia a rifampicina de *Mtb* tienen mutaciones genéticas en el gen *rpoB* codificador de la  $\beta$ -subunidad de la ARN polimerasa. Estos producen cambios conformacionales de la proteína que reducen la afinidad para el fármaco y en el desarrollo de la droga resistencia (Telenti et al., 1993).

En aproximadamente el 96% de aislados de *Mtb* resistentes a la rifampicina, existen mutaciones en la denominada "región de punto caliente" de 81-bp codones que abarcan 507-533 del gen *rpoB*. Esta sección es conocida como la región determinante de resistencia a rifampicina (S. Ramaswamy & Musser, 1998). Las mutaciones genéticas en los codones 516, 526 y 531 son las mutaciones más comúnmente asociados con la resistencia a la rifampicina en

la mayoría de los estudios (P. Coll et al., 2005; A. Somoskovi, L. M. Parsons, & M. Salfinger, 2001). Aunque con reducida frecuencia, se han observado identificado mutaciones externas a la región de puntos calientes de *rpoB* (Heep, Rieger, Beck, & Lehn, 2000; Siu et al., 2011). Las mutaciones en codones 518 o 529 se han asociado con un reducido nivel de resistencia a la rifampicina (Cavusoglu, Karaca-Derici, & Bilgic, 2004; Yang et al., 1998). La monorresistencia a la rifampicina es bastante rara, puesto que generalmente las cepas resistentes a rifampicina también son resistentes a otras drogas, especialmente a la isoniazida. Por ellos, se considera que la resistencia a la rifampicina es un marcador sustituto para MDR-TB (Traore, Fissette, Bastian, Devleeschouwer, & Portaels, 2000).

Los analisis de la secuenciación del genoma de *Mtb* han mostrado la presencia de mutaciones compensatorias en *rpoA* y *rpoC*, que codifican las subunidades  $\alpha$  y  $\beta$  de la enzima ARN polimerasa, en cepas resistentes a rifampicina con mutaciones genéticas en *rpoB* (Comas et al., 2012). Las mutaciones compensatorias serían las responsables de la restablecimiento de la capacidad biológica de estas cepas in-vivo y se han asociado con una transmisibilidad más frecuente (Brandis & Hughes, 2013; de Vos et al., 2013).

## **5.2 ISONIAZIDA**

La isoniazida es un agente anti-TB y junto con rifampicina, es utilizada como la base para el tratamiento de la enfermedad de tuberculosis. Isoniazida trabaja como un pro-fármaco que necesita la activación y transformacion de la enzima

catalasa/peroxidasa que es codificada por KatG, para ejercer el efecto (Y. Zhang, Heym, Allen, Young, & Cole, 1992). La Isoniazida inhibe la síntesis de un tipo de ácido micólico mediante una proteína portadora enoil-acilo dependiente de NADH (ACP) reductasa, que es codificada por *inhA* (Rawat, Whitty, & Tonge, 2003). La isoniazida es activa contra los bacilos en fase de replicación. La resistencia a esta droga se ha asociado a mutaciones en los genes *inhA*, *katG*, *ahpC* y *kasA*.

Existen dos mecanismos moleculares principales de resistencia a la droga isoniazida que se asocian con mutaciones en *katG*, *inhA* o en la región promotora. Diferentes estudios han encontrado mutaciones en estos genes más frecuentemente asociados a la resistencia de isoniazida (Hazbon et al., 2006; S. V. Ramaswamy et al., 2003). La mutación más frecuente ha sido identificada en S315T del gen *katG* que resulta en un producto de isoniazida deficiente en su conformación estructural necesaria para ejercer su actividad antimicrobiana (Rozwarski, Grant, Barton, Jacobs, & Sacchettini, 1998; Vilcheze & Jacobs, 2007). Esta mutación se ha correlacionado con resistencia de alto nivel (MIC>1mg/ml) a isoniazida (Fenner et al., 2012) y ocurre más frecuentemente en las cepas resistentes a varios tipos de fármacos (Hazbon et al., 2006). La segunda mutación más frecuente se identifica en la región promotora de *inhA* lo que origina una sobreexpresión de InhA o una mutación en su sitio activo, lo que podría disminuir su afinidad (Rozwarski et al., 1998). La mutación más frecuente encontrado es en la posición -15C/T y es más comúnmente asociado con el bajo nivel de resistencia a la isoniazida (MIC<1mg/ml). Las mutaciones genéticas en *inh* causan resistencia a la

isoniazida y también a etionamida, fármaco estructuralmente asociado, que comparte el mismo objetivo (Banerjee et al., 1994; Larsen et al., 2002). Un reporte encontró que la mutación en la región reguladora *inh* y en la región codificante *inh* causó alto nivel de resistencia a la isoniazida y resistencia a la etionamida (Machado et al., 2013).

En *Mtb*, una hidropoxidasa reductasa codificada por el gen *ahpC*, está asociada con la resistencia a los compuestos intermedios de oxígeno reactivo y que sus mutaciones en el promotor de *ahpC* son del tipo compensatorias por la reducida actividad catalasa y peroxidasa y no a causa de su resistencia a la isoniazida (Sherman et al., 1996) .

### 5.3 ETAMBUTOL

El etambutol forma parte del régimen de primera línea actual para el tratamiento de la tuberculosis. El etambutol es bacteriostático interfiere contra la multiplicación de los bacilos para la biosíntesis de arabinogalactano de la pared celular (Takayama & Kilburn, 1989). En *Mtb*, los genes *embCAB*, que son organizados en un operón alcanzan a codificar una arabinosil transferasa, que está asociada en la síntesis de arabinogalactano (Mikusova, Slayden, Besra, & Brennan, 1995).

El mecanismo de resistencia al etambutol se ha asociado con mutaciones en el gen *embB* en la posición *embB306* como el más frecuente en la mayoría de los estudios científicos (Sreevatsan et al., 1997; Telenti et al., 1997). Algunos estudios, sin embargo, también han identificado mutaciones en *embB306* en cepas que son sensibles a etambutol (Ahmad, Jaber, & Mokaddas, 2007). Las mutaciones en *embB306* pueden causar un grado variable de droga resistencia a etambutol y son requeridos, pero no son suficientes para originar un alto nivel de droga resistencia a etambutol. La presencia de un remanente de 30% de cepas que son resistentes a etambutol y que no presentan ninguna mutación en *embB* se convierte en una necesidad de identificar otros tipos posibles de mecanismos de resistencia a nivel del genoma a los medicamentos a etambutol.

## 5.4 PIRAZINAMIDA

La pirazinamida constituye ahora parte del régimen estándar de primera línea para el tratamiento de la tuberculosis. Pirazinamida es un análogo de la nicotinamida y su introducción comercial permitió la reducción de la duración del tratamiento contra la tuberculosis a seis meses. La pirazinamida tiene la propiedad de inhibir bacilos semi-inactivos que residen en ambientes ácidos, como los que se encuentran en las lesiones de tuberculosis (Mitchison, 1985). La pirazinamida es un pro-fármaco que debe ser convertido a una forma activa, el ácido pirazinoico, por la enzima pirazinamidasa/nicotinamidasa codificada por el gen *pncA* de Mtb (Konno, Feldmann, & McDermott, 1967; Scorpio & Zhang, 1996). El mecanismo de acción implica la conversión de pirazinamida al ácido pirazinoico, la que interrumpe la membrana de Mtb la cual inhibe el transporte de membrana. La droga pirazinamida es capaz de ingresar a la célula bacteriana por el mecanismo de difusión pasiva y después se convierte a ácido pirazinoico para ser excretado por una bomba de flujo débil. En condiciones ácidas, el ácido pirazinoico es reabsorbido en la célula y se acumula en el interior celular, debido a una bomba de salida, lo que produce un daño celular (Y. Zhang & Mitchison, 2003).

La existencia de mutaciones en el gen *pncA* permanecen como el descubrimiento más frecuente en cepas resistentes pirazinamida. Estas mutaciones, están dispersos por todo el gen, pero la mayoría de estas se producen en la región 561pb de marco de lectura abierto o en la región de 82pb de su promotor putativo (Jureen, Werngren, Toro, & Hoffner, 2008; Scorpio et al., 1997). Algunos pocos estudios han informado de la aparición de cepas



resistentes pirazinamida sin ninguna mutación en *pncA*, lo que indica que la resistencia podría ser debida a mutaciones en otro gen regulador aún no identificado (S. J. Cheng, Thibert, Sanchez, Heifets, & Zhang, 2000).

## 5.5 ESTREPTOMICINA

La estreptomicina fue aislada inicialmente del suelo a partir de la bacteria *Streptomyces griseus*, la estreptomicina fue el primer antibiótico para ser utilizado con éxito contra la tuberculosis. Desafortunadamente, tan pronto como se prescribió, la resistencia empezó, esto debido a que fue administrado como monoterapia (Crofton & Mitchison, 1948). La estreptomicina es un tipo de glicósido activo de aminociclitol frente a bacilos en etapa de crecimiento activo y su mecanismo de acción es por intermedio de la inhibición del inicio de la traducción para la síntesis de proteínas (Moazed & Noller, 1987). Estreptomicina tiene el mecanismo de actuar a nivel de la subunidad 30S de los ribosomas en el 16S rRNA y ribosomal 12S las cuales son codificadas por los genes *rrs* y *rpsL*, respectivamente (Finken, Kirschner, Meier, Wrede, & Bottger, 1993).

Las mutaciones en *rrs* y *rpsL* son los principales mecanismos que ejercen resistencia a la estreptomicina, sin embargo solo representan entre el 60%-70% de la resistencia corroborada (Gillespie, 2002). Entre las mutaciones genéticas encontradas en *rpsL*, la sustitución en el codón 43 de lisina por arginina es la más frecuente. Este tipo de mutaciones produce altos niveles de

resistencia a estreptomicina. El gen *rrs* tiene mutaciones más frecuentes que se producen entre los nucleótidos 530 y 915.

Recientemente, también se ha encontrado que las mutaciones en el gen *gidB*, codifica una metiltransferasa conservada para la subunidad 16S rRNA, este tipo origina resistencia de niveles bajos a estreptomicina (Okamoto et al., 2007; Spies, da Silva, Ribeiro, Rossetti, & Zaha, 2008).

## **5.6 FLUOROQUINOLONAS**

Las fluoroquinolonas se utilizan en la actualidad como fármacos de segunda línea para el tratamiento MDR-TB. La ciprofloxacina y la ofloxacina son componentes de derivados sintéticos del ácido nalidíxico (Goss, Deitz, & Cook, 1965). Las quinolonas de nueva generación, que incluye moxifloxacino y gatifloxacino que continúan siendo evaluados en ensayos clínicos y se plantean como los antibióticos de primera línea con la finalidad de reducir la duración del tratamiento antituberculosis (Palomino & Martin, 2013; Rustomjee et al., 2008).

El mecanismo de acción de las fluoroquinolonas es por un tipo de inhibición de la topoisomerasa IV y la topoisomerasa II (ADN girasa) dos enzimas clave para la viabilidad bacteriana. Las enzimas están codificadas por los genes *parC*, *parE*, *gyrA* y *gyrB*, respectivamente (Fabrega, Madurga, Giralt, & Vila, 2009). La topoisomerasa de tipo II es un tetrámero obtenido por dos subunidades  $\alpha$  y  $\beta$ , codificadas por los genes *gyrA* y *gyrB*, respectivamente, que tiene función

de catalizar el superenrollamiento del ADN (Takiff et al., 1994). La topoisomerasa de tipo II (ADN girasa) tiene como objetivo la actividad de las fluoroquinolonas (Aubry, Pan, Fisher, Jarlier, & Cambau, 2004). El principal mecanismo de desarrollo de resistencia a fluoroquinolonas en *Mtb* es por mutaciones cromosómicas en la región de resistencia a quinolona determinada en *gyrA* o *gyrB*. Las mutaciones más prevalentes encontradas son determinadas en la posición 90 y 94 del gen *gyrA* sin embargo, mutaciones genéticas en las posición 74, 88 y 91 también se han identificado (A. F. Cheng et al., 2004; Sun et al., 2008).

Un descubrimiento interesante en *Mtb* es la presencia de un polimorfismo genético natural en la posición 95 de gen *gyrA* la cual no se encuentra relacionado con resistencia a las fluoroquinolonas puesto que también se ha identificado en cepas susceptibles a fluoroquinolona (Musser, 1995). Una investigación interesante ha sido la ocurrencia simultánea de mutaciones A90G y T80A en el gen *gyrA* que causa hipersusceptibilidad a diferentes quinolonas (Aubry et al., 2006). Estas evidencias podrían señalar que el problema de la resistencia a fluoroquinolonas en *Mtb* podría tener un mecanismo más complejo de lo que inicialmente se pensaba en sus inicios.

### **5.7 KANAMICINA, CAPREOMICINA, AMIKACINA, VIOMICINA**

Los cuatro antibióticos son las drogas de segunda línea utilizados para el tratamiento de MDR-TB. Los cuatro antibióticos comparten el mecanismo de acción a través de la inhibición de la síntesis de proteínas, mientras

capreomicina y viomicina son antibióticos peptídicos cíclicos, la kanamicina y amikacina son aminoglicósidos.

Kanamicina y amikacina pueden inhibir la síntesis de proteínas por modificación estructural a nivel de 16S rRNA. Las mutaciones más frecuentes que se encuentran en las cepas resistentes a la kanamicina son en la posición 1400 y 1401 del gen *rrs*. Sin embargo, también se ha informado de mutaciones en la posición 1483 (Alangaden et al., 1998; Suzuki et al., 1998). Diferentes estudios han demostrado que los patrones de resistencia y sus niveles variables sugieren mecanismos de resistencia distintos que podrían ser posible (Kruuner, Jureen, Levina, Ghebremichael, & Hoffner, 2003). En relacion con esto, un bajo nivel de resistencia a kanamicina se ha relacionado con mutaciones en una región promotora del gen *eis*, la cual codifica un aminoglucósido acetiltransferasa (Zaunbrecher, Sikes, Metchock, Shinnick, & Posey, 2009). Mutaciones en la posición -10 y -35 del promotor de EIS condujo a una sobreexpresión de la proteína y de bajo nivel de resistencia a la kanamicina pero no a la amikacina. Estas mutaciones se encontraron en el 80% de los aislados clínicos que muestran bajo nivel de resistencia a la kanamicina (Campbell et al., 2011; Zaunbrecher et al., 2009).

Capreomicina y viomicina tienen una estructura análoga y se pueden unir en el mismo sitio del ribosoma, en las interfases de las subunidades pequeñas y grandes del ribosoma (Stanley, Blaha, Grodzicki, Strickler, & Steitz, 2010). Estas pueden mostrar resistencia cruzada según las investigaciones de estudios anteriores (McClatchy, Kanes, Davidson, & Moulding, 1977).

Las mutaciones identificadas en el gen *tlyA* también se han relacionado con la resistencia a capreomicina y viomicina. TlyA es una ARNr metiltransferasa específica para la metilación de ribosa en el rRNA. Las mutaciones en *tlyA* determinan que la actividad de metilación sea reducida (Johansen, Maus, Plikaytis, & Douthwaite, 2006).

### **5.8 ETIONAMIDA**

Etionamida es un componente derivado de ácido isonicotínico, que tiene una estructura similar a la isoniazida. Etionamida es un profármaco que requiere la activación de una monooxigenasa que la codifica el gen *ethA* (Morlock, Metchock, Sikes, Crawford, & Cooksey, 2003). Esta puede interferir con la síntesis de ácidos micólicos por la formación de un complejo con NAD la cual inhibe la enzima enoil-ACP reductasa. La resistencia a etionamida se puede producir por mutaciones en *etaA/ethA*, *ethR*, así también mutaciones en *inhA*, pueden causar resistencia a etionamida y sobre isoniazida (Brossier, Veziris, Truffot-Pernot, Jarlier, & Sougakoff, 2011; DeBarber, Mdluli, Bosman, Bekker, & Barry, 2000).

### **5.9 ÁCIDO PARAAMINOSALICÍLICO**

Acido Paraaminosalicilico fue uno de los primeros medicamentos antituberculosos usados en el tratamiento de la tuberculosis, junto con la estreptomicina y la isoniazida, el ácido paraaminosalicílico o por sus siglas PAS

se considera uno de los medicamentos de segunda línea del régimen de tratamiento para TB-MDR. En la actualidad, su mecanismo de acción no ha logrado ser completamente definido. Se ha propuesto que el ser un análogo del ácido benzoico para-amino, debe competir con él por la dihidropteroato sintasa, interfiriendo en el proceso de la síntesis de ácido fólico. Una investigación que utilizó mutagenesis de transposones para determinar las mutaciones en el gen *thyA* relacionada con la resistencia a PAS estuvieron también presentes en los aislados clínicos resistentes para PAS (Rengarajan et al., 2004). Una investigación también ha determinado varias mutaciones sin sentido en *folC* la cual codifica dihidrofolato sintasa la cual confiere resistencia a PAS en aislamientos de *Mtb* (Zhao et al., 2014). En un conjunto de 85 aislados clínicos de TB-MDR, las mutaciones en *folC* se identificaron en cinco cepas resistentes a PAS. No obstante, un 40% aproximadamente de las cepas resistentes a PAS tenían mutaciones en *thyA* lo que prueba que exista todavía otros mecanismos de resistencia (Mathys et al., 2009).

## **5.10 CICLOSERINA**

La cicloserina es un medicamento oral bacteriostático de segunda línea contra la enfermedad de la tuberculosis la cual es utilizado en regímenes de tratamiento de la TB-MDR. Es un análogo de alanina que, por intermedio del bloqueo de su actividad, d-alanina ligasa inhibe la síntesis de peptidoglicano. Este fármaco puede inhibir d-alanina racemasa que son necesarios para la conversión metabólica de l-alanina a d-alanina (Y. Zhang, 2005). Sin embargo,

el objetivo real del fármaco cicloserina en *Mtb* aun no está completamente estudiado, sin embargo en estudios de *M. smegmatis* se manifestó que la sobreexpresión de d-alanina racemasa logro la resistencia a cicloserina en los mutantes recombinantes (N. E. Caceres et al., 1997). Actualmente, se ha demostrado que la mutación puntual en el gen *cycA*, fue parcialmente responsable de causar resistencia a la cicloserina en *Mycobacterium bovis* BCG (Chen, Uplekar, Gordon, & Cole, 2012).

#### **5.11 TIOACETAZONA**

Tioacetazona es un farmaco antiguo utilizado para el tratamiento de la tuberculosis por su favorable actividad in vitro contra *M. tuberculosis* y por tener un costo muy reducido. Tiene inconvenientes de toxicidad, sin embargo, solo en pacientes co-infectados con VIH. Pertenece al grupo de 5 fármacos de la OMS y actúa inhibiendo la síntesis de ácidos micólicos (Grzegorzewicz et al., 2012).

#### **5.12 MACRÓLIDOS**

Los macrolidos son comúnmente recomendadas para el tratamiento de infecciones micobacterianas por su actividad limitada contra *Mtb*. La resistencia a los macrólidos se ha atribuido a la reducida permeabilidad de su pared celular y la expresión de *emr37*, un gen que codifica para una metilasa

a un sitio específico de 23S rRNA, bloqueando la unión del antibiótico. En investigaciones realizadas en *M. tuberculosis* y *M. microti* se determinó que la resistencia intrínseca fue inductible con concentraciones sub-inhibitorias de claritromicina, lo que conlleva a un aumento de hasta ocho veces los valores de CMI (Andini & Nash, 2006).

La resistencia a drogas en tuberculosis es el resultado de mutaciones genéticas en *Mtb* y se atribuye que la principal razón de bacterias resistentes es el incumplimiento de los regímenes de tratamiento. El tratamiento estándar de tuberculosis exige seis meses continuos de cuatro fármacos, en el caso de tuberculosis MDR se extiende por 18-24 meses e involucra drogas de segunda línea. Por lo que el cumplimiento de los regímenes de tratamiento contra tuberculosis puede ser un reto. El rol de la genómica en el estudio en la tuberculosis ha incrementado el entendimiento y descubrimiento de nuevas mutaciones, esto está permitiendo conocer más sobre el origen de la drogorresistencia y cómo se podría afrontar mejor.

## **6. La genómica de *Mycobacterium tuberculosis***

El secuenciamiento genómico de todo el genoma (WGS, por sus siglas en inglés, *Whole Genome Sequencing*) basado en tecnología de secuenciamiento de siguiente generación (NGS, por sus siglas en inglés, *Next Generation Sequencing*), ofrece nuevas oportunidades en la investigación y la salud pública, incluyendo el diagnóstico de laboratorio de TB y la epidemiología molecular (Nikolayevskyy, Kranzer, Niemann, & Drobniewski, 2016). NGS permite la secuenciación de genomas completos de múltiples cepas de forma simultánea, lo que simplifica el trabajo y proporciona mucha



más información en comparación a los métodos tradicionales (Walker et al., 2015). WGS permite identificar varios polimorfismos mediante comparaciones de referencia incluyendo SNP potencialmente importantes para la epidemiología, diagnóstico, detección de la resistencia a los medicamentos y las investigaciones filogenéticas (Loman et al., 2012; Niemann & Supply, 2014). En los últimos años se ha dado un incremento la capacidad de secuenciamiento genómico y mejorar la disponibilidad en laboratorios de rutina en los países de altos ingresos económicos. El genoma de *Mtb* tiene una tasa de mutación polimórfica de SNP de 0.5 por genoma por año (T. M. Walker et al., 2013). Es importante destacar que los datos obtenidos por WGS son de doble uso tanto a nivel de identificación de linaje como también para proporcionar información crucial sobre las mutaciones asociadas a resistencia de la tuberculosis. La secuencia de ADN de *Mycobacterium tuberculosis* H37Rv, ha sido investigada para mejorar la comprensión de la genética de este patógeno que tiene la característica de crecimiento lento y para ayudar a la concepción de nuevas intervenciones terapéuticas y profilácticas. El genoma completo comprende 4'411'529pb y esta integrado por alrededor de 4'000 genes, característicos por tener un alto contenido de guanina y citosina que se comprueba en los aminoácidos contenidos en las proteínas. *M. tuberculosis* difiere radicalmente de otras bacterias porque una gran parte de su codificación se diferencia a la producción de enzimas implicadas en la lipólisis, lipogénesis y a proteínas ricas en glicina que presentan estructura repetitiva, que se asocia a una fuente de variación antigénica (Cole et al., 1998).

En el mundo se han reportado como cepas prevalentes los genotipos siguientes: Latino América-Mediterráneo (LAM), Haarlem (H), Este de África-India (EAI), T, Central-Asia (CAS), Beijing y X. Los nombres están atribuidos a los lugares de procedencia donde fue identificado inicialmente (Philip Supply et al., 2006). Así

mismo, las recientes investigaciones en análisis de genomas han identificado regiones nuevas de *Mtb* de potenciales polimorfismos en asociación a resistencia de drogas antituberculosas (Casali et al., 2012; Ford et al., 2013; Timothy M Walker et al., 2013).

En el Perú los pobladores nativos también fueron infectados por TB (Culqui et al., 2010), pero estos aun permaneces desconocidos. Las familias genéticas Beijing de *M. tuberculosis* aparecieron desde el ingreso de migrantes asiáticos (Iwamoto et al., 2012).

La diversidad genética de *Mtb* en el Perú de aislados sensible y drogoresistentes revela principalmente a genotipos frecuentes: T (23.8%), LAM (23.8%), Haarlem (22.3%) y Beijing (9.3%)(Taype et al., 2012). Sin embargo, Grandjean *et al.*, en una investigación que involucro 2,139 aislados de *Mtb*, determinó que la familia LAM representa como minimo el 50% de los casos en el Peru (Grandjean et al., 2015).

Los genomas completos de *Mtb* sensible (Tarazona, Borda, Galarza, Agapito, & Guio, 2014), MDR (Galarza, Tarazona, Borda, Agapito, & Guio, 2014), XDR (Guio, Levano, Sanchez, & Tarazona, 2015) aislados en Perú, están resultando en información valiosa para determinar la variaciones genéticas, genes asociados a drogo resistencia que pueden ser frecuentes en el Peru. Las cepas de *Mtb* con resistencia a los fármacos de este patógeno que posiblemente son derivadas por el uso excesivo de antibióticos han planteado desde hace tiempo serias amenazas para la salud pública en todo el mundo. Es posible obtener un espectro más detallado de los mecanismos de resistencia a los medicamentos a nivel genómico utilizando la metodología genómica comparativa a una gran escala. Se ha encontrado que dos aislados de la familia de Beijing estrechamente vinculados, uno resistente a cuatro fármacos de

primera línea (CCDC5180) y uno sensible a ellos (CCDC5079), se secuenciaron su ADN por completo (Y. Zhang et al., 2011).

Una investigación secuenció el ADN de tres genomas de aislados de *Mtb* de procedencia rusa de diferentes fenotipos: susceptibles a fármacos, MDR y XDR las cuales son endémicas para Rusia. Los aislados fueron de diferentes familias genéticas, Beijing, LAM y Ural, lo que generó problemas en la comparación directa de sus genomas. Por ello, realizaron su comparación en el contexto más general de todos los genomas de *M. tuberculosis* disponibles en la base de datos del GenBank. El conjunto de SNP únicos que son no sinónimos para cada aislado secuenciado se integro por comparación de todos los SNP identificados dentro del mismo grupo filogenético. Un análisis funcional identificó que todas las proteínas con SNP únicos se asociaron a 20 clases funcionales diferentes grupos de genes ortólogos (COG). Probablemente los SNP únicos descubiertos en la cepa XDR pueda tener un rol importante en la adaptación de este microorganismo a su entorno y en el escape del tratamiento con medicamentos antituberculosos, lo que podría correlacionarse con las mutaciones encontradas en genes asociados a estadios de división celular como el gen *para*, incluso relacionados a biosíntesis de lipopolisacáridos y lípidos evidenciado en mutaciones de los genes *pks1* y *pks7*, *pks12* (Iliina et al., 2013).

## 7. La genómica de *Mycobacterium tuberculosis* para el diagnóstico

Recientemente la genómica ha tomado un rol importante en las investigaciones, incluso en aplicaciones en la práctica clínica. Sin embargo, la utilidad del secuenciamiento genómico de todo el genoma de *Mtb* como rutina en laboratorio aún permanece limitada (Witney et al., 2016). WGS en *Mtb*, se basa en ADN específico de un cultivo bacteriano puro, en el que ya habría sido diagnosticado el paciente con tuberculosis. Recientemente se han realizado estudios sobre la viabilidad de la secuenciación genómica de *Mtb* directamente de la muestra clínica (Brown et al., 2015; Doughty, Sergeant, Adetifa, Antonio, & Pallen, 2014), estos resultados no han sido muy alentadores, pues la cobertura es muy baja (0.002-0.7x), con un mapeo de lecturas de secuenciamiento al genoma humano superior al genoma de *Mtb* (20-99%), sin embargo un enfoque de enriquecimiento con oligonucleótidos específicos permite mejorar la cobertura (20x).

Existe ventajas en usar el WGS ya que es una técnica mas rápida, pero es necesario mejorar la identificación sobre la muestra de rutina de Tuberculosis, ya que la gran mayoría de muestras son negativas para *Mtb* (Demers et al., 2012). Los estudios proporcionan una prueba del principio que utilizando WGS de *Mtb* puede ser factible, pero un costo de secuenciamiento (2,500.0 soles, Celera Technologies), el requerimiento de expertos técnicos, equipamiento y la necesidad de bioinformáticos para convertir archivos de secuencias a reporte de laboratorio la convierte en un gran reto para el diagnóstico rutinario. Un logro más acorde a la realidad sería la amplificación de ácido nucleicos u otras técnicas (Galarza et al., 2016) de genes asociados a resistencia de drogas, para lograr una identificación más rápida.

Mediante el WGS, en teoría podemos conocer todas las mutaciones asociadas a resistencia para poder guiar el tratamiento clínico utilizando bases de datos especializadas (Flandrois, Lina, & Dumitrescu, 2014; Sandgren et al., 2009). En la práctica esto funcionaría si, estos marcadores predicen con exactitud la resistencia fenotípica *in vitro*. Un estudio más reciente basado en WGS y las concordancias del genotipo y fenotipo para las principales drogas asociadas a droga resistencia alcanza una limitada sensibilidad RIF (96.2%), INH (92.8%), PZA (70.9%), para otras drogas como EMB, STR, AMK, CAP, ETH, KAN, MOX, OFX tuvo menos que 81.7%(F. Coll et al., 2015). Por lo tanto, si WGS sustituye a la prueba fenotípica un grupo de personas recibirían los fármacos potencialmente hepatotóxicos (Guio et al., 2015), con poco o ningún efecto. Validar la metodología en la práctica clínica aún requiere más investigación que apoye la determinación exacta de SNP asociados a drogaresistencia.

### **III. HIPÓTESIS Y OBJETIVOS**

#### **1. Hipótesis**

El aislado sensible de *M. tuberculosis* comparte mayor información genética con los aislados MDR que con el aislado XDR de *M. tuberculosis*.

#### **2. Objetivo General**

Analizar con herramientas bioinformáticas la composición genética de tres genomas de *M. tuberculosis*, sensible, MDR y XDR

##### **a. Objetivos específicos**

- Determinar la estructura genómica de *M. tuberculosis* en tres aislados peruanos.
- Comparar secuencias genómicas de *M. tuberculosis* sensibles y drogo resistentes.

## IV. MATERIAL Y MÉTODOS

### 1. CEPAS, SENSIBILIDAD A DROGAS Y PURIFICACIÓN DE ADN

Las cepas se cultivaron en el medio Löwenstein-Jensen y los estudios microbiológicos de susceptibilidad a drogas fue por (1) método estándar Agar Proporciones en Placa realizado en el Laboratorio de Referencia Nacional de Micobacterias del Centro Nacional de Salud Pública (CNSP) del Instituto Nacional de Salud (INS), luego (2) la extracción del ADN genómico se realizó mediante el kit PureLink (Invitrogen, US) y (3) el proceso de cuantificación fue en un NanoDrop 8000 (Thermo Scientific, US) del Laboratorio de Referencia Nacional de Biotecnología y Biología Molecular CNSP. Las características de las cepas de MTB sensible, multidrogorresistente y extremadamente resistente de la investigación, se describen en la Tabla 1. Utilizando el procesamiento por MIRU-VNTR (Mycobacterial Interspersed Repetitive Unit-Variable Number Of Tandem Repeat) se determinó el tipo de linaje.

**Tabla 1.** Descripción de las cepas peruanas de *Mycobacterium tuberculosis* que forman parte de este estudio

Cepa	Fenotipo	Genotipo (SNPs)	MIRU-VNTR*
INS-SEN	RIF <sup>s</sup> , INH <sup>s</sup> , EMB <sup>s</sup> , PZA <sup>s</sup> , ETH <sup>s</sup> , AMI <sup>s</sup> , CAP <sup>s</sup> , OFL <sup>s</sup>	-	46[14]337645758468357578756
INS-MDR	RIF <sup>r</sup> , INH <sup>r</sup> , EMB <sup>r</sup> , PZA <sup>r</sup> , ETH <sup>r</sup> , AMI <sup>r</sup> , CAP <sup>r</sup> , OFL <sup>r</sup>	<i>rpoB</i> (D516V), <i>katG</i> (S315T), <i>kasA</i> (G269S), <i>thyA</i> (T202A), <i>pncA</i> (Q10R)	13[11]224342225237264433522
INS-XDR	RIF <sup>r</sup> , INH <sup>r</sup> , EMB <sup>r</sup> , PZA <sup>r</sup> , ETH <sup>r</sup> , KAN <sup>r</sup> , CAP <sup>r</sup> , OFL <sup>r</sup>	<i>rpoB</i> (D516V), <i>katG</i> (S315T), <i>kasA</i> (G269S), <i>gyrA</i> (D94G), <i>thyA</i> (T202A), <i>pncA</i> (Q10R), <i>embB</i> (Y319S)	23[11]225353335238264523533

RIF: rifampicina INH: isoniácida, EMB: etambutol, PZA: pirazinamida, ETH: etionamida, KAN: kanamicina, CAP: capreomicina, OFL: ofloxacina.

\*24-VNTR: MIRU02, Mtub04, ETRC, MIRU04, MIRU40, MIRU10, MIRU16, Mtub21, MIRU20, QUB11b, ETRA, Mtub29, Mtub30, ETRB, MIRU23, MIRU24, MIRU26, MIRU27, Mtub34, MIRU31, Mtub39, QUB26, QUB4156 y MIRU39.

### 2. Datos genómicos de cepas de *Mycobacterium tuberculosis*

Para el estudio se elaboraron librerías genómicas utilizando las secuencias de los extremos de pares, luego procesadas por método de

secuenciamiento en un equipo HiSeq 2000, Illumina (Macrogen Inc, Korea) (15-17), mediante la recomendación del fabricante. El proceso de ensamblaje de las lecturas de secuenciamiento fue llevado a cabo mediante el programa BWA v 0.5.8-r16 tomando como base el genoma referencia *Mycobacterium tuberculosis* H37Rv.

### **3. Linaje y secuenciamiento genómico**

El ADN genómico de INS-SEN se secuenció a una cobertura de 1,406x, que consistió en 61,422,158 lecturas de pares de bases (Tarazona et al., 2014), mientras que INS-MDR se secuencio a una cobertura de 1.331x con 58,157,302 lecturas (Galarza et al., 2014) y finalmente INS-XDR se secuencio a 1,140x produciendo 49,793,402 (Guio, Tarazona, Galarza, Borda, & Curitomay, 2014), utilizando la máquina secuenciadora Illumina HiSeq 2000. La determinación del linaje de esta cepa se basó en 24 unidades repetitivas micobacterianas interespaciadas – por polimorfismos de un solo nucleótido (SNP) basados en filogenia (Filliol et al., 2006) y número variable de repeticiones en tandem (MIRU-VNTR) (P. Supply et al., 2006).

### **4. Ensamblaje y anotación genómica**

Las secuencias genómicas INS-SEN, INS-MDR y INS-XDR se ensamblaron con BWA v 0.5.9-r16 (Li & Durbin, 2009), usando el genoma H37Rv (AL123456.3) como referencia, produciendo 18 contigs, 22 contigos y 20 contigs. La anotación genómica de las secuencias que son codificantes en los genomas INS-SEN, INS-MDR y INS-XDR fue mediante los programas RAST (*Rapid Annotations using Subsystem Technology*) (Aziz et al., 2008; Overbeek et al., 2014) y PGAAP (*Prokaryotic Genome*



*Annotation Pipeline*)(Tatusova et al., 2016), que son plataformas que utilizan Glimmer2 (Delcher, Harmon, Kasif, White, & Salzberg, 1999) para la identificación de genes, tRNAscan (Lowe & Eddy, 1997) para la identificación de tRNA y rRNA. Incluyendo las bases de datos FIGfams (Meyer, Overbeek, & Rodriguez, 2009) y proteínas NCBI (Klimke et al., 2009).

## **5. Identificación de SNP, COG y análisis comparativo**

Se determinó los SNPs asociados de *Mtb* H37Rv (AL123456.2) utilizando los software SamTools (Li et al., 2009), SNPtree (Leekitcharoenphon et al., 2012) y scripts en Perl (Stajich et al., 2002) para INS-SEN (Tarazona et al., 2014), INS-MDR (Galarza et al., 2014) e INS-XDR (Guio et al., 2014). Luego las proteínas fueron agrupadas en clúster de grupos ortólogos (COG)(Galperin, Kristensen, Makarova, Wolf, & Koonin, 2017) mediante BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990; Mount, 2007) y finalmente los SNP organizados en COG.

Un análisis genómico comparativo se realizó utilizando el criterio de inclusión/exclusión, una técnica de conteo que obtiene el numero de elementos en la unión de dos conjuntos finitos, donde los elementos de doble recuento son aquellos en la intersección de los tres conjuntos y el recuento se corrige restando los elementos de la intersección, para nuestro caso corresponde usar un total de 785, 805 y 814 de SNPs de los genomas INS-SEN, INS-MDR y INS-XDR, respectivamente, utilizando las bases de datos en Linux.

Se emplearon genomas adicionales de *M. tuberculosis* XDR que incluyen FJ05194, GuangZ0019 (Lin, Liu, Zhou, Wang, & Fleming, 2013), XDR1219, XDR1221 (Wu et al., 2013), GuangZ0008, 293, FJ07070, GuangZ0017, GuangZ0026 (H. Zhang et al., 2013); Perú, Perú\_04\_R0292, Perú\_03\_R0268 (Farhat et al., 2013); Rusia, CTRI-4 (Ilin et al., 2013), Rusia\_03-R1082 (Farhat et al., 2013); Malasia, UM1072388579 (Ng et al., 2013), y Sudáfrica, KZN\_R506 (Ioerger et al., 2009) y KZN\_605 (Z. L. Cui, Wang, Wang, Lu, & Hu, 2010). En INS-XDR se realizó un secuenciamiento Sanger para corroborar las mutaciones de los genes de resistencia en *rrs* y promotor *eis*.

Luego se realizó un estudio comparativo de polimorfismos de un solo nucleótido (SNPs) de las cepas de los genomas INS-SEN, INS-MDR y INS-XDR utilizando el criterio de inclusión/exclusión (Tarazona, D. et al. 2016).

## RESULTADOS

### 1. ESTRUCTURA DE LOS GENOMAS DE *M. tuberculosis* DE AISLADOS PERUANOS

El genomas INS-SEN, INS-MDR y INS-XDR tuvieron similares características en las lecturas de secuencias. Pero INS-SEN tuvo mayor numero de lecturas y en contraste mayor cobertura del genoma. Los dos aislados INS-SEN y INS-MDR tuvieron el mismo tamaño del genoma y INS-XDR tuvo mas tamaño. En cuanto al contenido de GC las tres cepas mostraron características estrechamente similares, mas detalles en la Tabla 2.

Tabla 2. Características de los genomas de *Mycobacterium tuberculosis*

	INS-SEN	INS-MDR	INS-XDR
<b>NCBI</b>	JAQH01000000	JAQI010000000	JANH01000000
<b>lecturas de secuencias</b>	61'422'158	58'157'302	49'793'402
<b>cobertura del genoma</b>	1406x	1331x	1140x
<b>tamaño</b>	4'383'671 pb	4'383'671 pb	4'391'020 pb
<b>contenido GC</b>	65,6%	65,6%	65,4%.

En el análisis genetico, se encontró que las tres secuencias genómicas INS-SEN, INS-MDR y INS-XDR contienen 4'389 secuencias codificantes, 45 tRNA y 03 tRNA

(Figura 2). El estudio comparativo genómico respecto a H37Rv determino que los genomas INS-SEN, INS-MDR y INS-XDR presentan 499, 805 y 815 SNPs, respectivamente. Esto permitio determinar los porcentajes de SNPs identificados en las regiones codificantes para los genomas INS-SEN con 88,2%, INS-MDR con 87,3% y INS-XDR con 87,3%, aproximadamente.

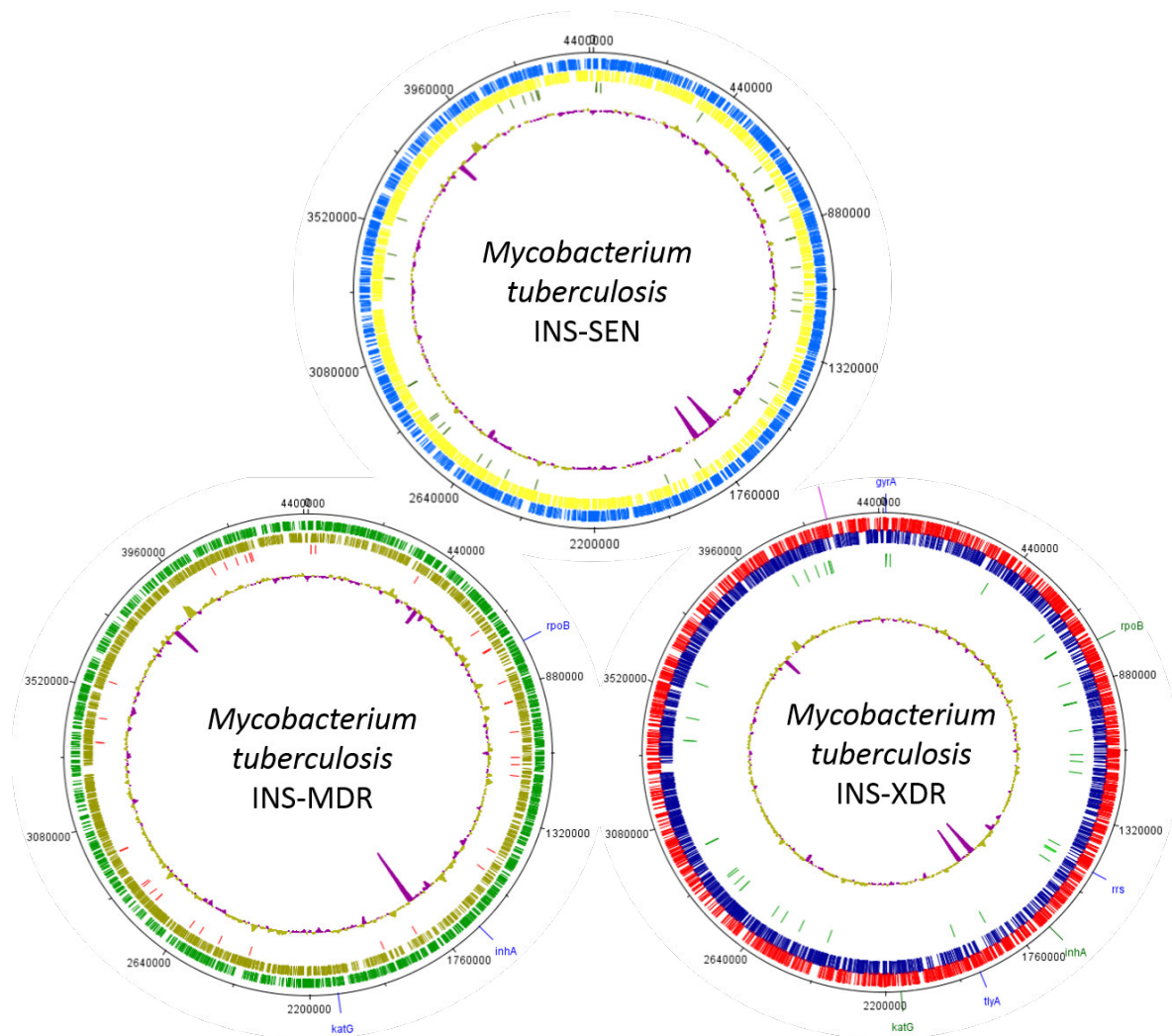


Figura 2. Estructura genómica de *Mycobacterium tuberculosis* de aislados peruanos, la anotación genómica de *Mycobacterium tuberculosis* fue realizada con RAST

## 2. COMPARACIÓN DE SECUENCIAS DE *M. tuberculosis* EN AISLADOS PERUANOS

El estudio comparativo de secuencias genómicas de *Mtb* del genotipo LAM utilizando el criterio de inclusión/exclusión generó que los genomas INS-SEN, INS-MDR y INS-XDR contienen SNP compartidos y específicos (Figura 3). Los números de SNPs específicos del genoma INS-SEN fue de 253, INS-MDR fue 20 y INS-XDR fue 24. Los conjuntos de SNP compartidos de los tres genomas equivale al 48,3% (525/1088). Los SNPs compartidos de las secuencias genómicas de INS-SEN y INS-MDR resultó en 26,2%; INS-SEN y INS-XDR resultó en 26,4%. Mientras que entre INS-MDR y INS-XDR comparten 93,9%.

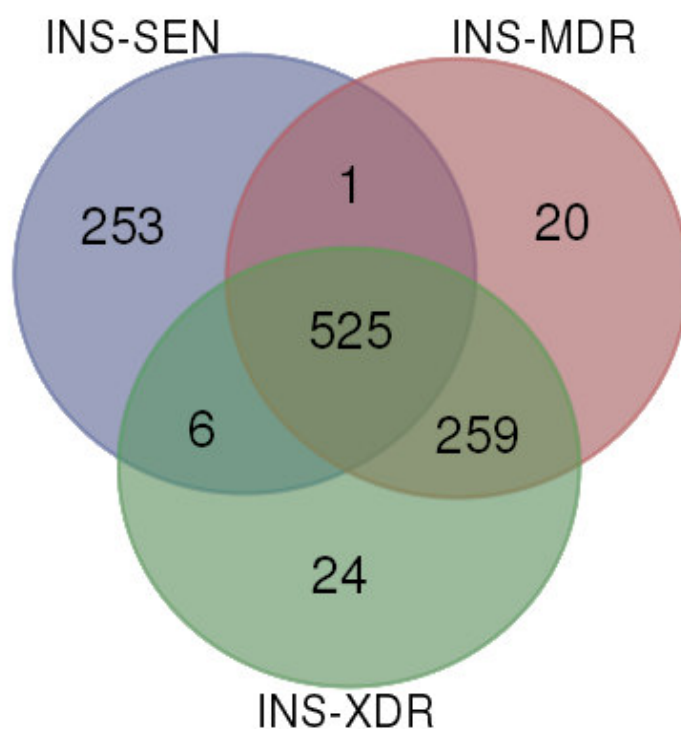


Figura 3. Diagrama de Venn de SNPs de las cepas de linaje LAM peruanas

### 3. ANÁLISIS DE GRUPOS ORTÓLOGOS (COGS)

En el genoma INS-SEN se identificó un total de 499 SNPs, 440 de ellos ubicados en las regiones codificantes del genoma, clasificados en las categorías COG: biosíntesis, transporte y catabolismo de metabolitos secundarios (Q) (n=38); transporte y metabolismo de lípidos (I) (n=35); replicación, recombinación y reparación (L) (n = 34); producción y conversión de energía (C) (n=32); transporte y metabolismo de aminoácidos (E) (n=31); metabolismo y transporte de carbohidratos (G) (n=27); motilidad celular (N) (n=26); biogénesis de la membrana (M) (n=24); metabolismo y transporte de la coenzima (H) (n=23); mecanismo de transducción de señales (T) (n=21); metabolismo y transporte de iones inorgánicos (P) (n=21); transcripción (K) (n=17); biogénesis, traducción y estructura ribosomal (n = 14); modificación postraduccional, recambio proteico y chaperonas (n = 13); transporte y metabolismo de nucleótidos (n = 12); mecanismos de defensa (n = 7); control del ciclo celular, división celular y partición cromosómica (n = 7); Procesamiento y modificación del ARN (n = 2); y tráfico, secreción y transporte de vesículas intracelulares (n = 2). No se identificaron genes mutaciones asociadas a resistencia. El genoma INS-MDR tiene un total identificado de 805 SNPs con respecto a H37Rv, de los cuales 703 se identificaron en regiones codificantes que lograron ser clasificados en las categorías COG de la siguiente manera: biosíntesis, catabolismo y transporte de metabolitos secundarios (n=38); metabolismo y transporte de lípidos y (n=41); recombinación, replicación y reparación (n=32); producción y conversión de energía (n=35); metabolismo y transporte de aminoácidos (n=30); metabolismo y transporte de carbohidratos (n=27); motilidad celular (n=21); biogénesis de la pared celular/membrana (M) (n=23); transporte y metabolismo de la coenzima (n=22); mecanismos de transducción de señales (n=22); transporte y metabolismo de iones

inorgánicos (n=22); transcripción (n=20); traducción, estructura y biogénesis ribosomal (n=16); modificación postraducciona l y recambio proteico (n=13); transporte de nucleótidos y metabolismo (n=11); mecanismos de defensa (n=5); control del ciclo celular, división celular y partición cromosómica (n=9); y procesamiento y modificación de ARN (n = 2). Se identifico una mutación genética en el gen *rpoB* (D516V) relacionada con la resistencia a la rifampina (Lipin, Stepanshina, Shemyakin, & Shinnick, 2007), mutaciones en los genes *katG* (S315T) y *kasA* (G269S) relacionadas con la resistencia a la isoniacida (Campbell et al., 2011; Lee, Lim, Tang, Telenti, & Wong, 1999) y una mutación en el gen *pncA* (Q10R) relacionado con la resistencia a la droga pirazinamida (S. J. Cheng et al., 2000). Además, se identifico mutaciones en el gen *gyrA* (E21Q, S95T, G247S, G668D) y una mutación en el gen *embB* (Y319S) que no se ha asociado a conferir resistencia (Lau et al., 2011; Malik, Willby, Sikes, Tsodikov, & Posey, 2012; Plinke, Rusch-Gerdes, & Niemann, 2006). En el genoma INS-XDR se identifico 218 SNPs no sinónimos (representando a 125 genes) principalmente en cuatro COG: (1) recombinación, replicación y reparación, L, n=14, (2) producción y conversión de energía, C, n=14, (3) metabolismo y transporte de aminoácidos, E, n=12 y (4) biosíntesis, transporte y catabolismo de metabolitos secundarios, Q, n=11.

Las secuencias genómicas en el presente incluyen: Sensible, MDR, XDR provenientes del genotipo LAM. Para ello, se construyo un listado de SNP individuales para cada secuencia genómica relativa a H37Rv. Los genes codificantes de proteínas que contienen SNP fueron relacionadas a veinte tipos de clases funcionales de COG. Se normalizaron los SNPs de *Mtb* utilizando de base el número total de SNPs identificados y genes para un determinado grupo de COG (Figura 4). Los grupos J: traducción, biogénesis ribosomal; K: transcripción; V: mecanismo de defensa; O:

modificación postraducciona, chaperonas; A: procesamiento y modificación de ARN y U: secreción, tráfico intracelular y transporte vesicular no tuvieron un arreglo estructural sobre su asignación, esto puede sugerir que genes de las presentes categorías estarían bajo presión evolutiva. Los genomas peruanos tienen número significativo de SNPs que se encuentran en proteínas L: replicación, recombinación y reparación y Q: transporte, biosíntesis y catabolismo de metabolismo secundario. En el grupo Q se identificó genes de síntesis de policétidos y *mce*. Mientras en el grupo L se identifican genes de fago phiRv1, ligasas, ADN girasa, metiltransferasas y transposasas, específicamente a las micobacterias con droga resistencia. El grupo COG H: transporte y metabolismo de coenzimas en la cepas de tipos sensible mostraron un elevado valor característico que cepas tipo MDR y XDR, las que tienen enzimas para síntesis de cobalamina, riboflavina, molibdeno, glutamina, magnesio quelatasa y nicotidamina. Así también el grupo G – transporte y metabolismo de carbohidratos tiene una significancia distinta con la secuencia genómica de la cepa sensible y las secuencias genómicas correspondientes a las cepas drogoresistentes, aquí encontramos a las glucosidasas transcetolasas, permeasas de la familia MFS, transportadores de azúcar y el gen *opcA*.



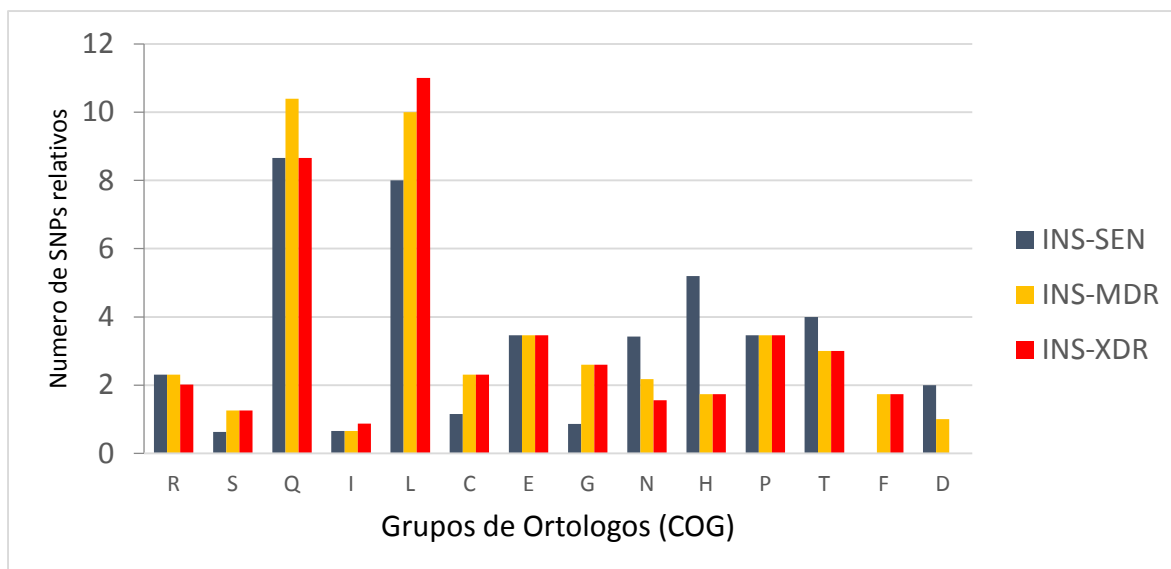


Figura 4. Distribución relativa de SNP encontrados en los genomas de MTB. C – producción y conversión de energía; D – control del ciclo celular, división celular; E – metabolismo y transporte de aminoácidos; F – metabolismo y transporte de nucleótido; G – metabolismo y transporte de carbohidratos; H – metabolismos y transporte de coenzima, I – metabolismo y transporte de lípidos; L - replicación, recombinación y reparación; M – biogénesis pared celular, membrana N – motilidad celular; P – metabolismo y transporte de ion inorgánico; Q – transporte, biosíntesis, y catabolismo de metabolismo secundario; R – funciones generales; S – función desconocida; T – mecanismos de transducción de señales.

## V. DISCUSIÓN

Los avances en genómica comparativa de agentes infecciosos de enfermedades humanas permiten conocer con alta precisión los polimorfismos asociados a la resistencia a drogas, factores de virulencia y procesos metabólicos brinda un valioso índice de impacto para la salud pública. Esta investigación reporta un análisis de las secuencias genómicas de *Mtb* aisladas de pacientes con procedencia de la ciudad de Lima, Perú, estas cepas son sensible, MDR y XDR. Estos organismos pueden representar una muestra modelo de las cepas sensibles y drogoresistentes circulando en el Perú.

Los aislados de genómicos INS-SEN, INS-MDR e INS-XDR, demuestran un conjunto de SNPs relacionados a grupos funcionales L y Q. El grupo Q asociado a proteínas de virulencia importante en *Mtb* que incluye a la proteína Mce, que tiene actividad de inducir el proceso de substrato en una fase de infección con tuberculosis (Stavrum et al., 2012), capacidad de supervivencia e intracelular y modificación del perfil lipídico (Chandolia et al., 2014). Las mutaciones funcionales en *Mce1* causan hipervirulencia, pero las mutaciones en *Mce2* y *Mce3* otorgan cepas atenuadas (Senaratne et al., 2008). En esta investigación se evidenció que la cepa INS-SEN tiene una variación genética en Ser270Asn del gen *Mce2D* que se asocia a su atenuación y virulencia. Así también, en las secuencias genómicas INS-MDR y INS-XDR se identificó una mutación en Val87Ala de *Mce1B* que pueden ser los determinantes de generar modificaciones lipídicas en la membrana.

En las proteínas del grupo L se encontraron a los fagos PhiRv1, con un sistema activo de escisión, integración que puede dar propiedades de cambiar de posición sobre los genomas de *Mtb* (Bibb & Hatfull, 2002). La secuencia genómica INS-SEN demostró

mayor variabilidad genética en el gen *PhiRv1* que en las secuencias genómicas drogorresistentes INS-MDR y INS-XDR. Todavía no se tiene evidencia de relaciones de sensibilidad y drogorresistencia, estudios de *Mtb* determinar que puede variar su genética con un ambiente en estrés y con ello mejorar la eficacia biológica (Fan, Abd Alla, & Xie, 2016), esto permite mejorar el entendimiento en aislados sensibles. Las secuencias genéticas de tipo transposasas, se encuentran asociadas a sucesos microevolutivos, esto genera cambios genéticos para la adaptación del microorganismo y su hospedero (Perez-Lago, Navarro, Herranz, Bouza, & Garcia-de-Viedma, 2013).

En el genoma INS-SEN se determinó que los SNPs más frecuentes en su distribución están en los grupos I, L y Q de la categoría COG, además en las secuencias codificantes de PE-PGRS y PPE. Las PE-PGRS están relacionadas a la evasión inmune y variación antigénica (Koh, Soh, & Seah, 2009), proteínas ricas en secuencias GC repetidas, lo que confiere un aumento de recombinación y en consecuencia aumentar la variación génica. Las secuencias codificantes PPE radica en tener un rol en relación a una variación antigénica del ADN (Zheng et al., 2008), como un mecanismo para modificar las proteínas de superficie de membrana para eludir la respuesta inmune. Las PE-PGRS y PPE cumplen un rol esencial en la virulencia de *Mtb* (Ahmed, Das, & Mukhopadhyay, 2015), esto se evidencia en cluster de polimórficos de aislados sensible y drogorresistentes de *Mtb*. La secuencia INS-SEN por ser una cepa sensible puede tener mayor superioridad por tener más virulencia sobre los aislados MDR o XDR (Smith et al., 2014). Se requieren mayores análisis con un número mayor de muestras para las consecuencias de las consecuencias de mutaciones en el ADN y la asociación con la patogenicidad y la virulencia.

Las secuencias genómicas MDR y XDR de *Mtb* tienen fragmentos con mutaciones adicionales asociadas a la resistencia de drogas, porque poseer un rol importante para la adaptación, supervivencia y proliferación. La secuencia INS-MDR de genotipo LAM contiene genes de resistencia a las drogas INH y RIF, con alta similaridad génica a la cepa KZN1435. Estudios generan evidencia que esto es posiblemente a una expansión del brote KZN en el Perú (Galarza et al., 2014) y la cepa INS-XDR con mutaciones que confieren drogoresistente a INH, RIF, PZA, KAN, CAP y tiene una alta asociación a cepa KZN 605 (Guio et al., 2014). Estos determinantes encontrados sugieren que INS-MDR y la cepa INS-XDR se encuentran estrechamente relacionados por una evaluación de la evidencia filogenética del mismo clado, una mínima distancia y alta similaridad de contenido de SNP, que incluye 51 nucleótidos diferentes en contraste al genoma de 4'411'532 pb en *Mtb*.

Las cepas de los genomas MDR y XDR analizadas muestran que las secuencias metiltransferasas presentan mayor variabilidad genica, y debido al papel para la reparación del ADN del genoma que puede proveer una estabilidad genética, a pesar de una hostil exposición de los componentes de los macrófagos (Miggiano et al., 2013). Las mutaciones en los genes asociados a la síntesis de nicotinamida, riboflavina, glutamina, molibdeno, cobalamina y magnesio están exclusivamente en los genomas MDR y XDR, por lo que se sugiere que el genoma *Mtb* y el cambio de eficacia biológica efectúan una presión para que existan modificaciones o selecciones génicas en los microorganismos de tipo drogoresistentes para procesar y sintetizar moléculas importantes para su modificación de alternativas en sus rutas metabólicas o funcionamiento normal (Borrell & Gagneux, 2009; Hughes & Brandis, 2013; Muller, Borrell, Rose, & Gagneux, 2013). Se requieren más análisis en la

determinación del rol importante de los SNPs en la dinámica de la genómica de *Mtb* de los aislados peruanos drogorresistentes.

## VI. CONCLUSIONES

- Los aislados peruanos de *M. tuberculosis* de la familia LAM de este estudio comparten un núcleo de 525 variantes genéticas.
- Los aislados peruanos de *M. tuberculosis* de la familia LAM de este estudio con sensibilidad a drogas y drogorresistentes (MDR y XDR) no se encuentran estrechamente relacionados.
- La cepa INS-SEN comparte variantes genéticas en el genoma en promedio 26.5% con las cepas INS-MDR y INS-XDR, mientras que INS-MDR y INS-XDR comparten 93.9%. La cepa INS-SEN comparte menores variantes genéticas en su genoma con la cepas INS-MDR y con INS-XDR.
- Se sugieren dos alternativas de origen para estos aislados peruanos de *M. tuberculosis* drogorresistentes de genotipo LAM. La primera, una cepa como INS-SEN requiere aproximadamente 233 mutaciones para tener capacidades de una cepa INS-MDR o una cepa INS-XDR. La segunda, INS-SEN y las cepas de drogorresistentes (INS-MDR, INS-XDR) pueden haber sido originados de brotes distintos.
- La cepa INS-SEN presenta mayor cantidad de variantes genéticas en las categorías COG relacionadas a biosíntesis, transporte y catabolismo de metabolitos secundarios (Q), transporte y metabolismo de lípidos (I) y replicación, recombinación y reparación (L) lo que podría mejorar su virulencia. Las cepas INS-MDR y INS-XDR presentan mayores mutaciones en genes relacionados a cambios lipídicos en su membrana y metiltransferasas y esto debido al rol de reparación de ADN de Mtb para proveer mayor estabilidad genética.

## VII. RECOMENDACIONES

Los datos genómicos de *Mycobacterium tuberculosis* han aumentado rápidamente, pero información sobre los lugares donde es más prevalente como Perú aún permanecen limitados, se hace necesario incrementar la información genómica en *M. tuberculosis* a fin de mejorar los análisis genómicos de aislados peruanos y puedan ser estadísticamente comprobados. Adicionalmente, para corroborar los datos es necesario completar los vacíos de la secuencia genómica de *Mycobacterium tuberculosis* usando tecnología de secuenciamiento capilar. Ara clarificar las conclusiones, es necesario aumentar el número de análisis de genomas de aislados peruanos.

## BIBLIOGRAFÍA

- Ahmad, S., Jaber, A. A., & Mokaddas, E. (2007). Frequency of embB codon 306 mutations in ethambutol-susceptible and -resistant clinical Mycobacterium tuberculosis isolates in Kuwait. *Tuberculosis (Edinb)*, 87(2), 123-129. doi: 10.1016/j.tube.2006.05.004
- Ahmed, A., Das, A., & Mukhopadhyay, S. (2015). Immunoregulatory functions and expression patterns of PE/PPE family members: Roles in pathogenicity and impact on anti-tuberculosis vaccine and drug design. *IUBMB Life*, 67(6), 414-427. doi: 10.1002/iub.1387
- Alangaden, G. J., Kreiswirth, B. N., Aouad, A., Khetarpal, M., Igno, F. R., Moghazeh, S. L., . . . Lerner, S. A. (1998). Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, 42(5), 1295-1297.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-410. doi: 10.1016/S0022-2836(05)80360-2
- Andini, N., & Nash, K. A. (2006). Intrinsic macrolide resistance of the Mycobacterium tuberculosis complex is inducible. *Antimicrob Agents Chemother*, 50(7), 2560-2562. doi: 10.1128/AAC.00264-06
- Arriaza, B. T., Salo, W., Aufderheide, A. C., & Holcomb, T. A. (1995). Pre-Columbian tuberculosis in northern Chile: molecular and skeletal evidence. *Am J Phys Anthropol*, 98(1), 37-45. doi: 10.1002/ajpa.1330980104
- Aubry, A., Pan, X. S., Fisher, L. M., Jarlier, V., & Cambau, E. (2004). Mycobacterium tuberculosis DNA gyrase: interaction with quinolones and correlation with antimycobacterial drug activity. *Antimicrob Agents Chemother*, 48(4), 1281-1288.
- Aubry, A., Veziris, N., Cambau, E., Truffot-Pernot, C., Jarlier, V., & Fisher, L. M. (2006). Novel gyrase mutations in quinolone-resistant and -hypersusceptible clinical isolates of Mycobacterium tuberculosis: functional analysis of mutant enzymes. *Antimicrob Agents Chemother*, 50(1), 104-112. doi: 10.1128/AAC.50.1.104-112.2006
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., . . . Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9, 75. doi: 10.1186/1471-2164-9-75
- Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., . . . Jacobs, W. R., Jr. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science*, 263(5144), 227-230.
- Bates, J. H., & Stead, W. W. (1993). The history of tuberculosis as a global epidemic. *Med Clin North Am*, 77(6), 1205-1217.
- Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., & Small, P. M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, 284(5419), 1520-1523.
- Bibb, L. A., & Hatfull, G. F. (2002). Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. *Mol Microbiol*, 45(6), 1515-1526.
- Blanchard, J. S. (1996). Molecular mechanisms of drug resistance in Mycobacterium tuberculosis. *Annu Rev Biochem*, 65, 215-239. doi: 10.1146/annurev.bi.65.070196.001243
- Borrell, S., & Gagneux, S. (2009). Infectiousness, reproductive fitness and evolution of drug-resistant Mycobacterium tuberculosis. *Int J Tuberc Lung Dis*, 13(12), 1456-1466.
- Brandis, G., & Hughes, D. (2013). Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *J Antimicrob Chemother*, 68(11), 2493-2497. doi: 10.1093/jac/dkt224
- Brossier, F., Veziris, N., Truffot-Pernot, C., Jarlier, V., & Sougakoff, W. (2011). Molecular investigation of resistance to the antituberculous drug ethionamide in multidrug-resistant clinical isolates of Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, 55(1), 355-360. doi: 10.1128/AAC.01030-10



- Brown, A. C., Bryant, J. M., Einer-Jensen, K., Holdstock, J., Houniet, D. T., Chan, J. Z., . . . Breuer, J. (2015). Rapid Whole-Genome Sequencing of Mycobacterium tuberculosis Isolates Directly from Clinical Samples. *J Clin Microbiol*, 53(7), 2230-2237. doi: 10.1128/JCM.00486-15
- Caceres, N. E., Harris, N. B., Wellehan, J. F., Feng, Z., Kapur, V., & Barletta, R. G. (1997). Overexpression of the D-alanine racemase gene confers resistance to D-cycloserine in Mycobacterium smegmatis. *J Bacteriol*, 179(16), 5046-5055.
- Caceres, O., Rastogi, N., Bartra, C., Couvin, D., Galarza, M., Asencios, L., & Mendoza-Ticona, A. (2014). Characterization of the genetic diversity of extensively-drug resistant Mycobacterium tuberculosis clinical isolates from pulmonary tuberculosis patients in Peru. *PLoS One*, 9(12), e112789. doi: 10.1371/journal.pone.0112789
- Campbell, P. J., Morlock, G. P., Sikes, R. D., Dalton, T. L., Metchock, B., Starks, A. M., . . . Posey, J. E. (2011). Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, 55(5), 2032-2041. doi: 10.1128/AAC.01550-10
- Casali, N., Nikolayevskyy, V., Balabanova, Y., Ignatyeva, O., Kontsevaya, I., Harris, S. R., . . . Hoffner, S. E. (2012). Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome research*, 22(4), 735-745.
- Cavusoglu, C., Karaca-Derici, Y., & Bilgic, A. (2004). In-vitro activity of rifabutin against rifampicin-resistant Mycobacterium tuberculosis isolates with known rpoB mutations. *Clin Microbiol Infect*, 10(7), 662-665. doi: 10.1111/j.1469-0691.2004.00917.x
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., . . . Barrell, B. G. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685), 537-544. doi: 10.1038/31159
- Coll, F., McNerney, R., Preston, M. D., Guerra-Assuncao, J. A., Warry, A., Hill-Cawthorne, G., . . . Clark, T. G. (2015). Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med*, 7(1), 51. doi: 10.1186/s13073-015-0164-0
- Coll, P., Aragon, L. M., Alcaide, F., Espasa, M., Garrigo, M., Gonzalez, J., . . . Salvado, M. (2005). Molecular analysis of isoniazid and rifampin resistance in Mycobacterium tuberculosis isolates recovered from Barcelona. *Microb Drug Resist*, 11(2), 107-114. doi: 10.1089/mdr.2005.11.107
- Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., . . . Gagneux, S. (2012). Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet*, 44(1), 106-110. doi: 10.1038/ng.1038
- Comas, I., Homolka, S., Niemann, S., & Gagneux, S. (2009). Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. *PLoS One*, 4(11), e7815. doi: 10.1371/journal.pone.0007815
- Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C., & Dye, C. (2003). The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med*, 163(9), 1009-1021. doi: 10.1001/archinte.163.9.1009
- Crofton, J., & Mitchison, D. A. (1948). Streptomycin resistance in pulmonary tuberculosis. *Br Med J*, 2(4588), 1009-1015.
- Crubezy, E., Legal, L., Fabas, G., Dabernat, H., & Ludes, B. (2006). Pathogeny of archaic mycobacteria at the emergence of urban life in Egypt (3400 BC). *Infect Genet Evol*, 6(1), 13-21. doi: 10.1016/j.meegid.2004.12.003
- Crubezy, E., Ludes, B., Poveda, J. D., Clayton, J., Crouau-Roy, B., & Montagnon, D. (1998). Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. *C R Acad Sci III*, 321(11), 941-951.
- Cui, Z., Wang, J., Lu, J., Huang, X., & Hu, Z. (2011). Association of mutation patterns in gyrA/B genes and ofloxacin resistance levels in Mycobacterium tuberculosis isolates from East China in 2009. *BMC Infect Dis*, 11, 78. doi: 10.1186/1471-2334-11-78

- Cui, Z. L., Wang, X. L., Wang, J., Lu, J. M., & Hu, Z. Y. (2010). [The effects of gene mutation related to drug resistance and drug efflux pump in extensively drug-resistant tuberculosis clinical isolates]. *Zhonghua Jie He He Hu Xi Za Zhi*, 33(7), 505-509.
- Culqui, D. R., Trujillo, O. V., Cueva, N., Aylas, R., Salaverry, O., & Bonilla, C. (2010). [Tuberculosis in the indigenous population of Peru 2008]. *Rev Peru Med Exp Salud Publica*, 27(1), 8-15.
- Chandolia, A., Rathor, N., Sharma, M., Saini, N. K., Sinha, R., Malhotra, P., . . . Bose, M. (2014). Functional analysis of mce4A gene of Mycobacterium tuberculosis H37Rv using antisense approach. *Microbiol Res*, 169(9-10), 780-787. doi: 10.1016/j.micres.2013.12.008
- Chen, J. M., Uplekar, S., Gordon, S. V., & Cole, S. T. (2012). A point mutation in cycA partially contributes to the D-cycloserine resistance trait of Mycobacterium bovis BCG vaccine strains. *PLoS One*, 7(8), e43467. doi: 10.1371/journal.pone.0043467
- Cheng, A. F., Yew, W. W., Chan, E. W., Chin, M. L., Hui, M. M., & Chan, R. C. (2004). Multiplex PCR amplicon conformation analysis for rapid detection of gyrA mutations in fluoroquinolone-resistant Mycobacterium tuberculosis clinical isolates. *Antimicrob Agents Chemother*, 48(2), 596-601.
- Cheng, S. J., Thibert, L., Sanchez, T., Heifets, L., & Zhang, Y. (2000). pncA mutations as a major mechanism of pyrazinamide resistance in Mycobacterium tuberculosis: spread of a monoresistant strain in Quebec, Canada. *Antimicrob Agents Chemother*, 44(3), 528-532.
- de Vos, M., Muller, B., Borrell, S., Black, P. A., van Helden, P. D., Warren, R. M., . . . Victor, T. C. (2013). Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. *Antimicrob Agents Chemother*, 57(2), 827-832. doi: 10.1128/AAC.01541-12
- Dean, A. S., Cox, H., & Zignol, M. (2017). Epidemiology of Drug-Resistant Tuberculosis. *Adv Exp Med Biol*, 1019, 209-220. doi: 10.1007/978-3-319-64371-7\_11
- DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L. G., & Barry, C. E., 3rd. (2000). Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*, 97(17), 9677-9682.
- Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*, 27(23), 4636-4641.
- Demers, A. M., Verver, S., Boulle, A., Warren, R., van Helden, P., Behr, M. A., & Coetzee, D. (2012). High yield of culture-based diagnosis in a TB-endemic setting. *BMC Infect Dis*, 12, 218. doi: 10.1186/1471-2334-12-218
- Dorman, S. E., & Chaisson, R. E. (2007). From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nat Med*, 13(3), 295-298. doi: 10.1038/nm0307-295
- Doughty, E. L., Sergeant, M. J., Adetifa, I., Antonio, M., & Pallen, M. J. (2014). Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ*, 2, e585. doi: 10.7717/peerj.585
- Duong, D. A., Nguyen, T. H., Nguyen, T. N., Dai, V. H., Dang, T. M., Vo, S. K., . . . Caws, M. (2009). Beijing genotype of Mycobacterium tuberculosis is significantly associated with high-level fluoroquinolone resistance in Vietnam. *Antimicrob Agents Chemother*, 53(11), 4835-4839. doi: 10.1128/AAC.00541-09
- Fabrega, A., Madurga, S., Giralt, E., & Vila, J. (2009). Mechanism of action of and resistance to quinolones. *Microb Biotechnol*, 2(1), 40-61. doi: 10.1111/j.1751-7915.2008.00063.x
- Fan, X., Abd Alla, A. A., & Xie, J. (2016). Distribution and function of prophage phiRv1 and phiRv2 among Mycobacterium tuberculosis complex. *J Biomol Struct Dyn*, 34(2), 233-238. doi: 10.1080/07391102.2015.1022602
- Farhat, M. R., Shapiro, B. J., Kieser, K. J., Sultana, R., Jacobson, K. R., Victor, T. C., . . . Murray, M. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nat Genet*, 45(10), 1183-1189. doi: 10.1038/ng.2747

- Fenner, L., Egger, M., Bodmer, T., Altpeter, E., Zwahlen, M., Jaton, K., . . . the Swiss Molecular Epidemiology of Tuberculosis Study, G. (2012). Effect of mutation and genetic background on drug resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, *56*(6), 3047-3053. doi: 10.1128/AAC.06460-11
- Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbon, M. H., Bobadilla del Valle, M., . . . Alland, D. (2006). Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol*, *188*(2), 759-772. doi: 10.1128/JB.188.2.759-772.2006
- Finken, M., Kirschner, P., Meier, A., Wrede, A., & Bottger, E. C. (1993). Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol Microbiol*, *9*(6), 1239-1246.
- Flandrois, J. P., Lina, G., & Dumitrescu, O. (2014). MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics*, *15*, 107. doi: 10.1186/1471-2105-15-107
- Ford, C. B., Shah, R. R., Maeda, M. K., Gagneux, S., Murray, M. B., Cohen, T., . . . Fortune, S. M. (2013). *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nature genetics*, *45*(7), 784-790.
- Galagan, J. E. (2014). Genomic insights into tuberculosis. *Nat Rev Genet*, *15*(5), 307-320. doi: 10.1038/nrg3664
- Galarza, M., Fasabi, M., Levano, K. S., Castillo, E., Barreda, N., Rodriguez, M., & Guio, H. (2016). High-resolution melting analysis for molecular detection of multidrug resistance tuberculosis in Peruvian isolates. *BMC Infect Dis*, *16*(1), 260. doi: 10.1186/s12879-016-1615-y
- Galarza, M., Tarazona, D., Borda, V., Agapito, J. C., & Guio, H. (2014). Evidence of Clonal Expansion in the Genome of a Multidrug-Resistant *Mycobacterium tuberculosis* Clinical Isolate from Peru. *Genome Announc*, *2*(1). doi: 10.1128/genomeA.00089-14
- Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2017). Microbial genome analysis: the COG approach. *Brief Bioinform*. doi: 10.1093/bib/bbx117
- Garaycochea, O., & Ticona, E. (2015). [Routes of public transport and the situation of tuberculosis in Lima, Peru]. *Rev Peru Med Exp Salud Publica*, *32*(1), 93-97.
- Garcia de Viedma, D., Chaves, F., & Inigo, J. (2006). New route of importation of *Mycobacterium tuberculosis* Beijing genotype. *Emerg Infect Dis*, *12*(1), 169-170. doi: 10.3201/eid1201.041214
- Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., . . . Hewinson, R. G. (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A*, *100*(13), 7877-7882. doi: 10.1073/pnas.1130426100
- Gillespie, S. H. (2002). Evolution of drug resistance in *Mycobacterium tuberculosis*: clinical and molecular perspective. *Antimicrob Agents Chemother*, *46*(2), 267-274.
- Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., & Cole, S. T. (1999). Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol*, *32*(3), 643-655.
- Goss, W. A., Deitz, W. H., & Cook, T. M. (1965). Mechanism of Action of Nalidixic Acid on *Escherichia Coli*. Inhibition of Deoxyribonucleic Acid Synthesis. *J Bacteriol*, *89*, 1068-1074.
- Grandjean, L., Iwamoto, T., Lithgow, A., Gilman, R. H., Arikawa, K., Nakanishi, N., . . . Moore, D. A. (2015). The Association between *Mycobacterium Tuberculosis* Genotype and Drug Resistance in Peru. *PLoS One*, *10*(5), e0126271. doi: 10.1371/journal.pone.0126271
- Grzegorzewicz, A. E., Kordulakova, J., Jones, V., Born, S. E., Belardinelli, J. M., Vaquie, A., . . . Jackson, M. (2012). A common mechanism of inhibition of the *Mycobacterium tuberculosis* mycolic

- acid biosynthetic pathway by isoxyl and thiacetazone. *J Biol Chem*, 287(46), 38434-38441. doi: 10.1074/jbc.M112.400994
- Guio, H., Levano, K. S., Sanchez, C., & Tarazona, D. (2015). [The role of pharmacogenomics in the tuberculosis treatment regime]. *Rev Peru Med Exp Salud Publica*, 32(4), 794-800.
- Guio, H., Tarazona, D., Galarza, M., Borda, V., & Curitomay, R. (2014). Genome analysis of 17 extensively drug-resistant strains reveals new potential mutations for resistance. *Genome Announc*, 2(4). doi: 10.1128/genomeA.00759-14
- Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., . . . Vincent, V. (2005). Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog*, 1(1), e5. doi: 10.1371/journal.ppat.0010005
- Hazbon, M. H., Brimacombe, M., Bobadilla del Valle, M., Cavatore, M., Guerrero, M. I., Varma-Basil, M., . . . Alland, D. (2006). Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 50(8), 2640-2649. doi: 10.1128/AAC.00112-06
- Heep, M., Rieger, U., Beck, D., & Lehn, N. (2000). Mutations in the beginning of the *rpoB* gene can induce resistance to rifamycins in both *Helicobacter pylori* and *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 44(4), 1075-1077.
- Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., . . . Spigelman, M. (2008). Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS One*, 3(10), e3426. doi: 10.1371/journal.pone.0003426
- Horna-Campos, O. J., Sanchez-Perez, H. J., Sanchez, I., Bedoya, A., & Martin, M. (2007). Public transportation and pulmonary tuberculosis, Lima, Peru. *Emerg Infect Dis*, 13(10), 1491-1493. doi: 10.3201/eid1310.060793
- Hughes, D., & Brandis, G. (2013). Rifampicin Resistance: Fitness Costs and the Significance of Compensatory Evolution. *Antibiotics (Basel)*, 2(2), 206-216. doi: 10.3390/antibiotics2020206
- Iilina, E. N., Shitikov, E. A., Ikryannikova, L. N., Alekseev, D. G., Kamashev, D. E., Malakhova, M. V., . . . Govorun, V. M. (2013). Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia. *PLoS One*, 8(2), e56577. doi: 10.1371/journal.pone.0056577
- Ioerger, T. R., Koo, S., No, E. G., Chen, X., Larsen, M. H., Jacobs, W. R., Jr., . . . Sacchettini, J. C. (2009). Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One*, 4(11), e7778. doi: 10.1371/journal.pone.0007778
- Iwamoto, T., Grandjean, L., Arikawa, K., Nakanishi, N., Caviedes, L., Coronel, J., . . . Gilman, R. H. (2012). Genetic diversity and transmission characteristics of Beijing family strains of *Mycobacterium tuberculosis* in Peru. *PLoS One*, 7(11), e49651. doi: 10.1371/journal.pone.0049651
- Johansen, S. K., Maus, C. E., Plikaytis, B. B., & Douthwaite, S. (2006). Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Mol Cell*, 23(2), 173-182. doi: 10.1016/j.molcel.2006.05.044
- Jureen, P., Werngren, J., Toro, J. C., & Hoffner, S. (2008). Pyrazinamide resistance and *pncA* gene mutations in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 52(5), 1852-1854. doi: 10.1128/AAC.00110-08
- Kapur, V., Whittam, T. S., & Musser, J. M. (1994). Is *Mycobacterium tuberculosis* 15,000 years old? *J Infect Dis*, 170(5), 1348-1349.
- Kato-Maeda, M., Metcalfe, J. Z., & Flores, L. (2011). Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. *Future Microbiol*, 6(2), 203-216. doi: 10.2217/fmb.10.165
- Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciuffo, S., Fedorov, B., . . . Tatusova, T. (2009). The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Res*, 37(Database issue), D216-223. doi: 10.1093/nar/gkn734

- Koh, K. W., Soh, S. E., & Seah, G. T. (2009). Strong antibody responses to Mycobacterium tuberculosis PE-PGRS62 protein are associated with latent and active tuberculosis. *Infect Immun*, 77(8), 3337-3343. doi: 10.1128/IAI.01175-08
- Konno, K., Feldmann, F. M., & McDermott, W. (1967). Pyrazinamide susceptibility and amidase activity of tubercle bacilli. *Am Rev Respir Dis*, 95(3), 461-469. doi: 10.1164/arrd.1967.95.3.461
- Kremer, K., van Soolingen, D., Frothingham, R., Haas, W. H., Hermans, P. W., Martin, C., . . . van Embden, J. D. (1999). Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol*, 37(8), 2607-2618.
- Kruuner, A., Jureen, P., Levina, K., Ghebremichael, S., & Hoffner, S. (2003). Discordant resistance to kanamycin and amikacin in drug-resistant Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, 47(9), 2971-2973.
- Labidi, A., & Thoen, C. O. (1989). Genetic relatedness among Mycobacterium tuberculosis and M. bovis. *Acta Leprol*, 7 Suppl 1, 217-221.
- Larsen, M. H., Vilcheze, C., Kremer, L., Besra, G. S., Parsons, L., Salfinger, M., . . . Jacobs, W. R., Jr. (2002). Overexpression of inhA, but not kasA, confers resistance to isoniazid and ethionamide in Mycobacterium smegmatis, M. bovis BCG and M. tuberculosis. *Mol Microbiol*, 46(2), 453-466.
- Lau, R. W., Ho, P. L., Kao, R. Y., Yew, W. W., Lau, T. C., Cheng, V. C., . . . Yam, W. C. (2011). Molecular characterization of fluoroquinolone resistance in Mycobacterium tuberculosis: functional analysis of gyrA mutation at position 74. *Antimicrob Agents Chemother*, 55(2), 608-614. doi: 10.1128/AAC.00920-10
- Lee, A. S., Lim, I. H., Tang, L. L., Telenti, A., & Wong, S. Y. (1999). Contribution of kasA analysis to detection of isoniazid-resistant Mycobacterium tuberculosis in Singapore. *Antimicrob Agents Chemother*, 43(8), 2087-2089.
- Leekitcharoenphon, P., Kaas, R. S., Thomsen, M. C., Friis, C., Rasmussen, S., & Aarestrup, F. M. (2012). snpTree--a web-server to identify and construct SNP trees from whole genome sequence data. *BMC Genomics*, 13 Suppl 7, S6. doi: 10.1186/1471-2164-13-S7-S6
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. doi: 10.1093/bioinformatics/btp352
- Lin, N., Liu, Z., Zhou, J., Wang, S., & Fleming, J. (2013). Draft genome sequences of two super-extensively drug-resistant isolates of Mycobacterium tuberculosis from China. *FEMS Microbiol Lett*, 347(2), 93-96. doi: 10.1111/1574-6968.12238
- Lipin, M. Y., Stepanshina, V. N., Shemyakin, I. G., & Shinnick, T. M. (2007). Association of specific mutations in katG, rpoB, rpsL and rrs genes with spoligotypes of multidrug-resistant Mycobacterium tuberculosis isolates in Russia. *Clin Microbiol Infect*, 13(6), 620-626. doi: 10.1111/j.1469-0691.2007.01711.x
- Loman, N. J., Constantinidou, C., Chan, J. Z., Halachev, M., Sergeant, M., Penn, C. W., . . . Pallen, M. J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol*, 10(9), 599-606. doi: 10.1038/nrmicro2850
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, 25(5), 955-964.
- Machado, D., Perdigao, J., Ramos, J., Couto, I., Portugal, I., Ritter, C., . . . Viveiros, M. (2013). High-level resistance to isoniazid and ethionamide in multidrug-resistant Mycobacterium tuberculosis of the Lisboa family is associated with inhA double mutations. *J Antimicrob Chemother*, 68(8), 1728-1732. doi: 10.1093/jac/dkt090

- Malik, S., Willby, M., Sikes, D., Tsodikov, O. V., & Posey, J. E. (2012). New insights into fluoroquinolone resistance in *Mycobacterium tuberculosis*: functional genetic analysis of *gyrA* and *gyrB* mutations. *PLoS One*, 7(6), e39754. doi: 10.1371/journal.pone.0039754
- Mathys, V., Wintjens, R., Lefevre, P., Bertout, J., Singhal, A., Kiass, M., . . . Bifani, P. (2009). Molecular genetics of para-aminosalicylic acid resistance in clinical isolates and spontaneous mutants of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 53(5), 2100-2109. doi: 10.1128/AAC.01197-08
- McClatchy, J. K., Kanes, W., Davidson, P. T., & Moulding, T. S. (1977). Cross-resistance in *M. tuberculosis* to kanamycin, capreomycin and viomycin. *Tubercle*, 58(1), 29-34.
- Meyer, F., Overbeek, R., & Rodriguez, A. (2009). FIGfams: yet another set of protein families. *Nucleic Acids Res*, 37(20), 6643-6654. doi: 10.1093/nar/gkp698
- Miggiano, R., Casazza, V., Garavaglia, S., Ciaramella, M., Perugino, G., Rizzi, M., & Rossi, F. (2013). Biochemical and structural studies of the *Mycobacterium tuberculosis* O6-methylguanine methyltransferase and mutated variants. *J Bacteriol*, 195(12), 2728-2736. doi: 10.1128/JB.02298-12
- Mikusova, K., Slayden, R. A., Besra, G. S., & Brennan, P. J. (1995). Biogenesis of the mycobacterial cell wall and the site of action of ethambutol. *Antimicrob Agents Chemother*, 39(11), 2484-2489.
- Mitchison, D. A. (1979). Basic mechanisms of chemotherapy. *Chest*, 76(6 Suppl), 771-781.
- Mitchison, D. A. (1985). The action of antituberculosis drugs in short-course chemotherapy. *Tubercle*, 66(3), 219-225.
- Moazed, D., & Noller, H. F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, 327(6121), 389-394. doi: 10.1038/327389a0
- Morlock, G. P., Metchock, B., Sikes, D., Crawford, J. T., & Cooksey, R. C. (2003). *ethA*, *inhA*, and *katG* loci of ethionamide-resistant clinical *Mycobacterium tuberculosis* isolates. *Antimicrob Agents Chemother*, 47(12), 3799-3805.
- Mortimer, T. D., & Pepperell, C. S. (2014). Genomic signatures of distributive conjugal transfer among mycobacteria. *Genome Biol Evol*, 6(9), 2489-2500. doi: 10.1093/gbe/evu175
- Mount, D. W. (2007). Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc*, 2007, pdb top17. doi: 10.1101/pdb.top17
- Muller, B., Borrell, S., Rose, G., & Gagneux, S. (2013). The heterogeneous evolution of multidrug-resistant *Mycobacterium tuberculosis*. *Trends Genet*, 29(3), 160-169. doi: 10.1016/j.tig.2012.11.005
- Murray, C. J., Ortblad, K. F., Guinovart, C., Lim, S. S., Wolock, T. M., Roberts, D. A., . . . Vos, T. (2014). Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 384(9947), 1005-1070. doi: 10.1016/S0140-6736(14)60844-8
- Musser, J. M. (1995). Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clin Microbiol Rev*, 8(4), 496-514.
- Nerlich, A. G., & Losch, S. (2009). Paleopathology of human tuberculosis and the potential role of climate. *Interdiscip Perspect Infect Dis*, 2009, 437187. doi: 10.1155/2009/437187
- Ng, K. P., Yew, S. M., Chan, C. L., Chong, J., Tang, S. N., Soo-Hoo, T. S., . . . Yee, W. Y. (2013). Draft Genome Sequence of the First Isolate of Extensively Drug-Resistant (XDR) *Mycobacterium tuberculosis* in Malaysia. *Genome Announc*, 1(1). doi: 10.1128/genomeA.00056-12
- Nguyen, L., & Pieters, J. (2009). Mycobacterial subversion of chemotherapeutic reagents and host defense tactics: challenges in tuberculosis drug development. *Annu Rev Pharmacol Toxicol*, 49, 427-453. doi: 10.1146/annurev-pharmtox-061008-103123
- Niemann, S., & Supply, P. (2014). Diversity and evolution of *Mycobacterium tuberculosis*: moving to whole-genome-based approaches. *Cold Spring Harb Perspect Med*, 4(12), a021188. doi: 10.1101/cshperspect.a021188

- Nikolayevskyy, V., Kranzer, K., Niemann, S., & Drobniewski, F. (2016). Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: A systematic review. *Tuberculosis (Edinb)*, 98, 77-85. doi: 10.1016/j.tube.2016.02.009
- Okamoto, S., Tamaru, A., Nakajima, C., Nishimura, K., Tanaka, Y., Tokuyama, S., . . . Ochi, K. (2007). Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol Microbiol*, 63(4), 1096-1106. doi: 10.1111/j.1365-2958.2006.05585.x
- Ormerod, L. P. (2005). Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment. *Br Med Bull*, 73-74, 17-24. doi: 10.1093/bmb/ldh047
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., . . . Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*, 42(Database issue), D206-214. doi: 10.1093/nar/gkt1226
- Palomino, J. C., & Martin, A. (2013). Tuberculosis clinical trial update and the current anti-tuberculosis drug portfolio. *Curr Med Chem*, 20(30), 3785-3796.
- Pantel, A., Petrella, S., Matrat, S., Brossier, F., Bastian, S., Reitter, D., . . . Aubry, A. (2011). DNA gyrase inhibition assays are necessary to demonstrate fluoroquinolone resistance secondary to gyrB mutations in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 55(10), 4524-4529. doi: 10.1128/AAC.00707-11
- Perez-Lago, L., Navarro, Y., Herranz, M., Bouza, E., & Garcia-de-Viedma, D. (2013). Genetic features shared by *Mycobacterium tuberculosis* strains involved in microevolution events. *Infect Genet Evol*, 16, 326-329. doi: 10.1016/j.meegid.2013.02.016
- Plinke, C., Rusch-Gerdes, S., & Niemann, S. (2006). Significance of mutations in embB codon 306 for prediction of ethambutol resistance in clinical *Mycobacterium tuberculosis* isolates. *Antimicrob Agents Chemother*, 50(5), 1900-1902. doi: 10.1128/AAC.50.5.1900-1902.2006
- Ramaswamy, S., & Musser, J. M. (1998). Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis*, 79(1), 3-29. doi: 10.1054/tuld.1998.0002
- Ramaswamy, S. V., Reich, R., Dou, S. J., Jasperse, L., Pan, X., Wanger, A., . . . Graviss, E. A. (2003). Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 47(4), 1241-1250.
- Rawat, R., Whitty, A., & Tonge, P. J. (2003). The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the *Mycobacterium tuberculosis* enoyl reductase: adduct affinity and drug resistance. *Proc Natl Acad Sci U S A*, 100(24), 13881-13886. doi: 10.1073/pnas.2235848100
- Rengarajan, J., Sasseti, C. M., Naroditskaya, V., Sloutsky, A., Bloom, B. R., & Rubin, E. J. (2004). The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria. *Mol Microbiol*, 53(1), 275-282. doi: 10.1111/j.1365-2958.2004.04120.x
- Rozwarski, D. A., Grant, G. A., Barton, D. H., Jacobs, W. R., Jr., & Sacchettini, J. C. (1998). Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*. *Science*, 279(5347), 98-102.
- Rustomjee, R., Lienhardt, C., Kanyok, T., Davies, G. R., Levin, J., Mthiyane, T., . . . Gatifloxacin for, T. B. s. t. (2008). A Phase II study of the sterilising activities of ofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. *Int J Tuberc Lung Dis*, 12(2), 128-138.
- Salo, W. L., Aufderheide, A. C., Buikstra, J., & Holcomb, T. A. (1994). Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A*, 91(6), 2091-2094.
- Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B. K., Church, G. M., & Murray, M. B. (2009). Tuberculosis drug resistance mutation database. *PLoS Med*, 6(2), e2. doi: 10.1371/journal.pmed.1000002
- Scorpio, A., Lindholm-Levy, P., Heifets, L., Gilman, R., Siddiqi, S., Cynamon, M., & Zhang, Y. (1997). Characterization of pncA mutations in pyrazinamide-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 41(3), 540-543.

- Scorpio, A., & Zhang, Y. (1996). Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med*, 2(6), 662-667.
- Senaratne, R. H., Sidders, B., Sequeira, P., Saunders, G., Dunphy, K., Marjanovic, O., . . . Riley, L. W. (2008). Mycobacterium tuberculosis strains disrupted in *mce3* and *mce4* operons are attenuated in mice. *J Med Microbiol*, 57(Pt 2), 164-170. doi: 10.1099/jmm.0.47454-0
- Sheen, P., Couvin, D., Grandjean, L., Zimic, M., Dominguez, M., Luna, G., . . . Moore, D. A. (2013). Genetic diversity of Mycobacterium tuberculosis in Peru and exploration of phylogenetic associations with drug resistance. *PLoS One*, 8(6), e65873. doi: 10.1371/journal.pone.0065873
- Sherman, D. R., Mdluli, K., Hickey, M. J., Arain, T. M., Morris, S. L., Barry, C. E., 3rd, & Stover, C. K. (1996). Compensatory *ahpC* gene expression in isoniazid-resistant Mycobacterium tuberculosis. *Science*, 272(5268), 1641-1643.
- Siu, G. K., Zhang, Y., Lau, T. C., Lau, R. W., Ho, P. L., Yew, W. W., . . . Yam, W. C. (2011). Mutations outside the rifampicin resistance-determining region associated with rifampicin resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother*, 66(4), 730-733. doi: 10.1093/jac/dkq519
- Smith, K. L., Saini, D., Bardarov, S., Larsen, M., Frothingham, R., Gandhi, N. R., . . . Lee, S. (2014). Reduced virulence of an extensively drug-resistant outbreak strain of Mycobacterium tuberculosis in a murine model. *PLoS One*, 9(4), e94953. doi: 10.1371/journal.pone.0094953
- Somaskovi, A., Parsons, L. M., & Salfinger, M. (2001). The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in Mycobacterium tuberculosis. *Respiratory research*, 2(3), 164.
- Somaskovi, A., Parsons, L. M., & Salfinger, M. (2001). The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in Mycobacterium tuberculosis. *Respir Res*, 2(3), 164-168.
- Soudani, A., Hadjfredj, S., Zribi, M., Messaoud, T., Masmoudi, A., Majed, B., & Fendri, C. (2010). First report of molecular characterization of fluoroquinolone-resistant Mycobacterium tuberculosis isolates from a Tunisian hospital. *Clin Microbiol Infect*, 16(9), 1454-1457. doi: 10.1111/j.1469-0691.2009.03087.x
- Spies, F. S., da Silva, P. E., Ribeiro, M. O., Rossetti, M. L., & Zaha, A. (2008). Identification of mutations related to streptomycin resistance in clinical isolates of Mycobacterium tuberculosis and possible involvement of efflux mechanism. *Antimicrob Agents Chemother*, 52(8), 2947-2949. doi: 10.1128/AAC.01570-07
- Sreevatsan, S., Stockbauer, K. E., Pan, X., Kreiswirth, B. N., Moghazeh, S. L., Jacobs, W. R., Jr., . . . Musser, J. M. (1997). Ethambutol resistance in Mycobacterium tuberculosis: critical role of *embB* mutations. *Antimicrob Agents Chemother*, 41(8), 1677-1681.
- Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., . . . Birney, E. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res*, 12(10), 1611-1618. doi: 10.1101/gr.361602
- Stanley, R. E., Blaha, G., Grodzicki, R. L., Strickler, M. D., & Steitz, T. A. (2010). The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nat Struct Mol Biol*, 17(3), 289-293. doi: 10.1038/nsmb.1755
- Stavrum, R., Valvatne, H., Stavrum, A. K., Riley, L. W., Ulvestad, E., Jonassen, I., . . . Grewal, H. M. (2012). Mycobacterium tuberculosis Mce1 protein complex initiates rapid induction of transcription of genes involved in substrate trafficking. *Genes Immun*, 13(6), 496-502. doi: 10.1038/gene.2012.24
- Sun, Z., Zhang, J., Zhang, X., Wang, S., Zhang, Y., & Li, C. (2008). Comparison of *gyrA* gene mutations between laboratory-selected ofloxacin-resistant Mycobacterium tuberculosis strains and clinical isolates. *Int J Antimicrob Agents*, 31(2), 115-121. doi: 10.1016/j.ijantimicag.2007.10.014
- Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., . . . Roring, S. (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-



- variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *Journal of clinical microbiology*, 44(12), 4498-4510.
- Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., . . . van Soolingen, D. (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol*, 44(12), 4498-4510. doi: 10.1128/JCM.01392-06
- Suzuki, Y., Katsukawa, C., Tamaru, A., Abe, C., Makino, M., Mizuguchi, Y., & Taniguchi, H. (1998). Detection of kanamycin-resistant *Mycobacterium tuberculosis* by identifying mutations in the 16S rRNA gene. *J Clin Microbiol*, 36(5), 1220-1225.
- Takayama, K., & Kilburn, J. O. (1989). Inhibition of synthesis of arabinogalactan by ethambutol in *Mycobacterium smegmatis*. *Antimicrob Agents Chemother*, 33(9), 1493-1499.
- Takiff, H. E., Salazar, L., Guerrero, C., Philipp, W., Huang, W. M., Kreiswirth, B., . . . Telenti, A. (1994). Cloning and nucleotide sequence of *Mycobacterium tuberculosis* *gyrA* and *gyrB* genes and detection of quinolone resistance mutations. *Antimicrob Agents Chemother*, 38(4), 773-780.
- Tarazona, D., Borda, V., Galarza, M., Agapito, J. C., & Guio, H. (2014). Functional Analysis Using Whole-Genome Sequencing of a Drug-Sensitive *Mycobacterium tuberculosis* Strain from Peru. *Genome Announc*, 2(1). doi: 10.1128/genomeA.00087-14
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., . . . Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*, 44(14), 6614-6624. doi: 10.1093/nar/gkw569
- Taype, C. A., Agapito, J. C., Accinelli, R. A., Espinoza, J. R., Godreuil, S., Goodman, S. J., . . . Shaw, M. A. (2012). Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Peru. *Infect Genet Evol*, 12(3), 577-585. doi: 10.1016/j.meegid.2012.02.002
- Telenti, A., Imboden, P., Marchesi, F., Lowrie, D., Cole, S., Colston, M. J., . . . Bodmer, T. (1993). Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet*, 341(8846), 647-650.
- Telenti, A., Philipp, W. J., Sreevatsan, S., Bernasconi, C., Stockbauer, K. E., Wieles, B., . . . Jacobs, W. R., Jr. (1997). The *emb* operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. *Nat Med*, 3(5), 567-570.
- Traore, H., Fissette, K., Bastian, I., Devleeschouwer, M., & Portaels, F. (2000). Detection of rifampicin resistance in *Mycobacterium tuberculosis* isolates from diverse countries by a commercial line probe assay as an initial indicator of multidrug resistance. *Int J Tuberc Lung Dis*, 4(5), 481-484.
- Udwadia, Z. F. (2012). Totally drug-resistant tuberculosis in India: who let the djinn out? *Respirology*, 17(5), 741-742. doi: 10.1111/j.1440-1843.2012.02192.x
- Valentina Alarcón, E. A., Cecilia Figueroa, Alberto Mendoza-Ticona. (2017). Tuberculosis en el Perú: situación epidemiológica, avances y desafíos para su control. *Revista Peruana de Medicina Experimental y Salud Publica*, 34(2), 299-310.
- Victor, T. C., de Haas, P. E., Jordaan, A. M., van der Spuy, G. D., Richardson, M., Van Soolingen, D., . . . Warren, R. (2004). Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a western cape F11 genotype. *Journal of clinical microbiology*, 42(2), 769-772.
- Vilcheze, C., & Jacobs, W. R., Jr. (2007). The mechanism of isoniazid killing: clarity through the scope of genetics. *Annu Rev Microbiol*, 61, 35-50. doi: 10.1146/annurev.micro.61.111606.122346
- Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., . . . Crook, D. W. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet infectious diseases*, 13(2), 137-146.
- Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., . . . Peto, T. E. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis*, 13(2), 137-146. doi: 10.1016/S1473-3099(12)70277-3

- Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Del Ojo Elias, C., Bradley, P., . . . Modernizing Medical Microbiology Informatics, G. (2015). Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis*, 15(10), 1193-1202. doi: 10.1016/S1473-3099(15)00062-6
- WHO. (2014). *Global Tuberculosis Report 2014*: World Health Organization.
- Witney, A. A., Cosgrove, C. A., Arnold, A., Hinds, J., Stoker, N. G., & Butcher, P. D. (2016). Clinical use of whole genome sequencing for *Mycobacterium tuberculosis*. *BMC Med*, 14, 46. doi: 10.1186/s12916-016-0598-2
- Wu, W., Zheng, H., Zhang, L., Wen, Z., Zhang, S., Pei, H., . . . Li, Y. (2013). A genome-wide analysis of multidrug-resistant and extensively drug-resistant strains of *Mycobacterium tuberculosis* Beijing genotype. *Mol Genet Genomics*, 288(9), 425-436. doi: 10.1007/s00438-013-0758-4
- Yang, B., Koga, H., Ohno, H., Ogawa, K., Fukuda, M., Hirakata, Y., . . . Kohno, S. (1998). Relationship between antimycobacterial activities of rifampicin, rifabutin and KRM-1648 and *rpoB* mutations of *Mycobacterium tuberculosis*. *J Antimicrob Chemother*, 42(5), 621-628.
- Zaunbrecher, M. A., Sikes, R. D., Jr., Metchock, B., Shinnick, T. M., & Posey, J. E. (2009). Overexpression of the chromosomally encoded aminoglycoside acetyltransferase *eis* confers kanamycin resistance in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*, 106(47), 20004-20009. doi: 10.1073/pnas.0907925106
- Zhang, H., Li, D., Zhao, L., Fleming, J., Lin, N., Wang, T., . . . Bi, L. (2013). Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet*, 45(10), 1255-1260. doi: 10.1038/ng.2735
- Zhang, Y. (2005). The magic bullets and tuberculosis drug targets. *Annu Rev Pharmacol Toxicol*, 45, 529-564. doi: 10.1146/annurev.pharmtox.45.120403.100120
- Zhang, Y., Chen, C., Liu, J., Deng, H., Pan, A., Zhang, L., . . . Wan, K. (2011). Complete genome sequences of *Mycobacterium tuberculosis* strains CCDC5079 and CCDC5080, which belong to the Beijing family. *J Bacteriol*, 193(19), 5591-5592. doi: 10.1128/JB.05452-11
- Zhang, Y., Heym, B., Allen, B., Young, D., & Cole, S. (1992). The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature*, 358(6387), 591-593. doi: 10.1038/358591a0
- Zhang, Y., & Mitchison, D. (2003). The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis*, 7(1), 6-21.
- Zhao, F., Wang, X. D., Erber, L. N., Luo, M., Guo, A. Z., Yang, S. S., . . . Deng, J. Y. (2014). Binding pocket alterations in dihydrofolate synthase confer resistance to para-aminosalicylic acid in clinical isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 58(3), 1479-1487. doi: 10.1128/AAC.01775-13
- Zheng, H., Lu, L., Wang, B., Pu, S., Zhang, X., Zhu, G., . . . Zhang, Y. (2008). Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PLoS One*, 3(6), e2375. doi: 10.1371/journal.pone.0002375
- Zink, A. R., Grabner, W., & Nerlich, A. G. (2005). Molecular identification of human tuberculosis in recent and historic bone tissue samples: The role of molecular techniques for the study of historic tuberculosis. *Am J Phys Anthropol*, 126(1), 32-47. doi: 10.1002/ajpa.10409

## ANEXO 1

Genome Announc. 2014 Feb 27;2(1). pii: e00087-14. doi: 10.1128/genomeA.00087-14.

# Functional Analysis Using Whole-Genome Sequencing of a Drug-Sensitive *Mycobacterium tuberculosis* Strain from Peru

D. Tarazona<sup>a</sup>, V. Borda<sup>a</sup>, M. Galarza<sup>a</sup>, J. C. Agapito<sup>c</sup>, H. Guio<sup>a,b</sup>

Author Affiliations

<sup>a</sup> Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud, Lima, Peru

<sup>b</sup> Asociación Latinoamericana de Biotecnología (ALBIOTEC), Lima, Peru

<sup>c</sup> Unidad de Biotecnología Molecular-Laboratorios de Investigación y Desarrollo de Ciência y Tecnología (LID), Universidad Peruana Cayetano Heredia, Lima, Peru

## ABSTRACT

We report the whole-genome sequence of a Latin American-Mediterranean (LAM) lineage drug-sensitive *Mycobacterium tuberculosis* strain from Peru, INS-SEN. The functional analysis revealed more mutations in secondary metabolite biosynthesis, transport, and catabolism (clusters of orthologous groups [COG] category Q) than for other LAM-sensitive strains. This study contributes to the understanding of the genomic diversity of drug-sensitive *M. tuberculosis*.

## GENOME ANNOUNCEMENT

In 2012, there were an estimated 8.6 million new cases of tuberculosis (TB) worldwide. In Peru, the incidence rate for TB was 95 cases/100,000 people, of which 96% of cases were drug-sensitive TB (1). It has been reported that in Peru there is a high diversity of *Mycobacterium tuberculosis* lineages, including Latin American-Mediterranean (LAM) (23.8%), Haarlem (23.8%), T (22.3%), and Beijing (9.3%) (3). We performed whole-genome sequencing and analysis to investigate the genetic diversity and phylogeny relationships of a drug-sensitive strain of *M. tuberculosis*, INS-SEN. INS-SEN was isolated from Lima, Peru. The establishment of this strain's lineage was based on 24 mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) loci (4) and by single-nucleotide polymorphisms (SNPs) based on phylogeny (5). The genomic DNA of INS-SEN was sequenced to 1,406× coverage, which consisted of 61,422,158 paired-end reads, using the Illumina HiSeq 2000 sequencer machine. Then, the genomic sequence was assembled with BWA v 0.5.9-r16 (6), using the H37Rv genome (AL123456.3) as a reference, producing 18 contigs. The genomic sequence was annotated with the Rapid Annotations using Subsystem Technology (RAST) server (7) and Prokaryotic Genome Annotation Pipeline (PGAAP). A polymorphism study of the INS-SEN genome was carried out by comparative analysis against the genome of the drug-sensitive strain KZN 4207 (LAM lineage) (8) using SNPsFinder (9) to identify the differences between intergenic and coding regions, and then clusters of orthologous groups (COG) (10).

The 24 loci for MIRU-VNTR and SNPs based on phylogeny determined that INS-SEN belongs to the LAM lineage. The genome sequence is about 99.98% completed compared to the H37Rv reference genome, which has a genome size of 4.42 Mb. The INS-SEN strain has a total of 4,383,671 bp, with an average GC content of 65.6%. It contains 4,389 predicted coding sequences (CDSs). A total of 499 polymorphisms were observed in our comparative study, with 440 of these located in the coding regions of the genome that were classified in the following COG categories: secondary metabolite biosynthesis, transport, and catabolism (Q) ( $n = 38$ ); lipid transport and metabolism (I) ( $n = 35$ ); replication, recombination, and repair (L) ( $n = 34$ ); energy production and conversion (C) ( $n = 32$ ); amino acid transport and metabolism (E) ( $n = 31$ ); carbohydrate transport and metabolism (G) ( $n = 27$ ); cell motility (N) ( $n = 26$ ); cell wall/membrane/envelope biogenesis (M) ( $n = 24$ ); coenzyme transport and metabolism (H) ( $n = 23$ ); signal transduction mechanisms (T) ( $n = 21$ ); inorganic ion transport and metabolism (P) ( $n = 21$ ); transcription (K) ( $n = 17$ ); translation, ribosomal structure, and biogenesis ( $n = 14$ ); posttranslational modification, protein turnover, and chaperones ( $n = 13$ ); nucleotide transport and metabolism ( $n = 12$ ); defense mechanisms ( $n = 7$ ); cell cycle control, cell division, and chromosome partitioning ( $n = 7$ ); RNA processing and modification ( $n = 2$ ); and intracellular trafficking, secretion, and vesicular transport ( $n = 2$ ).

INS-SEN had more SNPs in PPE associated with antigenic variation (11) in category N and in PE-PGRS associated with antigenic variation and immune evasion (12) in category M than the strains KZN 4207 and H37Rv. Additionally, INS-SEN showed more mutations in category Q than the strain KZN 4207. It is possible that the organization of SNPs in INS-SEN may have a role in adaptation to its environment.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [JAQH00000000](#). The version described in this paper is JAQH01000000.

## ACKNOWLEDGMENT

---

This study was supported by the Peruvian National Institute of Health.

## FOOTNOTES

---

Address correspondence to H. Guio, [heinnerguido@gmail.com](mailto:heinnerguido@gmail.com).

**Citation** Tarazona D, Borda V, Galarza M, Agapito JC, Guio H. 2014. Functional analysis using whole-genome sequencing of a drug-sensitive *Mycobacterium tuberculosis* strain from Peru. *Genome Announc.* 2(1):e00087-14. doi:10.1128/genomeA.00087-14.

**Received** 23 January 2014.

**Accepted** 3 February 2014.

**Published** 27 February 2014.

Copyright © 2014 Tarazona et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](#).

## REFERENCES

---

1. WHO. 2013. Global tuberculosis report 2013. World Health Organization, Geneva, Switzerland.
2. Reference deleted.
3. Taype CA, Agapito JC, Accinelli RA, Espinoza JR, Godreuil S, Goodman SJ, Bañuls AL, Shaw MA. 2012. Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Peru. *Infect. Genet. Evol.* 12:577–585. 10.1016/j.meegid.2012.02.002. CrossRefMedlineGoogle Scholar
4. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 44:4498–4510. 10.1128/JCM.01392-06. Abstract/FREE Full Text
5. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, Bobadilla del Valle M, Fyfe J, García-García L, Rastogi N, Sola C, Zozio T, Guerrero MI, León CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendón A, Sifuentes-Osornio J, Ponce de León A, Cave MD, Fleischmann R, Whittam TS, Alland D. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 188:759–772. 10.1128/JB.188.2.759-772.2006. Abstract/FREE Full Text
6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. 10.1093/bioinformatics/btp324. Abstract/FREE Full Text
7. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42:D206–D214. 10.1093/nar/gkt1226. Abstract/FREE Full Text
8. Iøerger TR, Koo S, No EG, Chen X, Larsen MH, Jacobs WR Jr, Pillay M, Sturm AW, Sacchettini JC. 2009. Genome analysis of multi- and extensively drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* 4:e7778. 10.1371/journal.pone.0007778. CrossRefMedlineGoogle Scholar
9. Song J, Xu Y, White S, Miller KW, Wolinsky M. 2005. SNPsFinder—a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. *Bioinformatics* 21:2083–2084. 10.1093/bioinformatics/bti176. Abstract/FREE Full Text
10. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. 10.1186/1471-2105-4-41. CrossRefMedlineGoogle Scholar

11. Zheng H, Lu L, Wang B, Pu S, Zhang X, Zhu G, Shi W, Zhang L, Wang H, Wang S, Zhao G, Zhang Y. 2008. Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PLoS One* 3:e2375. 10.1371/journal.pone.0002375. CrossRefMedlineGoogle Scholar
12. Koh KW, Soh SE, Seah GT. 2009. Strong antibody responses to *Mycobacterium tuberculosis* PE-PGRS62 protein are associated with latent and active tuberculosis. *Infect. Immun.* 77:3337–3343. 10.1128/IAI.01175-08.  
Abstract/FREE Full Text

Genome Announc. 2014 Feb 27;2(1). pii: e00089-14. doi: 10.1128/genomeA.00089-14.

## Evidence of Clonal Expansion in the Genome of a Multidrug-Resistant *Mycobacterium tuberculosis* Clinical Isolate from Peru

M. Galarza<sup>a</sup>, D. Tarazona<sup>a</sup>, V. Borda<sup>a</sup>, J. C. Agapito<sup>c</sup>, H. Guio<sup>a,b</sup>

Author Affiliations

<sup>a</sup> Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud, Lima, Peru

<sup>b</sup> Asociación Latinoamericana de Biotecnología (ALBIOTEC), Lima, Peru

<sup>c</sup> Unidad de Biotecnología Molecular-Laboratorios de Investigación y Desarrollo de Ciência y Tecnología (LID), Universidad Peruana Cayetano Heredia, Lima, Peru

### ABSTRACT

We report the genome sequence of *Mycobacterium tuberculosis* INS-MDR from Peru, a multidrug-resistant tuberculosis (MDR-TB) and Latin American-Mediterranean (LAM) lineage strain. Our analysis showed mutations related to drug resistance in the *rpoB* (D516V), *katG* (S315T), *kasA* (G269S), and *pncA* (Q10R) genes. Our evidence suggests that INS-MDR may be a clonal expansion related to the African strain KZN 1435.

### GENOME ANNOUNCEMENT

The emergence of multidrug-resistant (MDR) strains for tuberculosis (TB) is resulting in one of the major worldwide public health problems. In 2012, the incidence rate for TB in Peru was 95 cases/100,000 population, of which 3.9% were new cases of MDR-TB (1), and from these MDR-TB cases, 79% were from Lima (2). Currently, there is no genomic information on an MDR strain from Peru. Thus, obtaining this information is a key step in understanding the biology of the pathogen and improving treatment for TB (3). Here, we report genomic features of a multidrug-resistant strain of *Mycobacterium tuberculosis*, INS-MDR, from a patient with active TB from Peru.

The establishment of this strain's lineage was based on 24 mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) loci (4) and on single-nucleotide polymorphisms (SNPs) based on phylogeny (5). The genome sequencing of INS-MDR was performed using an Illumina HiSeq 2000 sequencer with coverage of 1,331×. Resulting paired-end reads were assembled with BWA v 0.5.9-r16 (6), using the H37Rv genome (NC000962.3) as a reference. The genomic sequence was annotated with the RAST server (7), Prokaryotic Genome Annotation Pipeline (PGAAP), and Clusters of Orthologous Groups (COG) (8) databases. A comparative analysis was carried out between INS-MDR and KZN 1435, an MDR strain from South Africa (9), using SNPsFinder (10) to identify the differences in intergenic coding regions and gene ontology classifications (COG).

It has been determined that INS-MDR belongs to the LAM lineage. We obtained 58,157,302 paired-end reads that, after the assembly, resulted in 22 contigs comprising about 99.98% compared to the H37Rv genome. INS-MDR is 4,383,671 bp long, with an average GC content of 65.6%.

A total of 805 polymorphisms with respect to H37Rv were observed, with 703 of these located in the coding regions that were classified in COG categories as follows: secondary metabolite biosynthesis, transport, and catabolism ( $n = 38$ ); lipid transport and metabolism ( $n = 41$ ); replication, recombination, and repair ( $n = 32$ ); energy production and conversion ( $n = 35$ ); amino acid transport and metabolism ( $n = 30$ ); carbohydrate transport and metabolism ( $n = 27$ ); cell motility ( $n = 21$ ); cell wall/membrane/envelope biogenesis (M) ( $n = 23$ ); coenzyme transport and metabolism ( $n = 22$ ); signal transduction mechanisms ( $n = 22$ ); inorganic ion transport and metabolism ( $n = 22$ ); transcription ( $n = 20$ ); translation, ribosomal structure, and biogenesis ( $n = 16$ ); posttranslational modification and protein turnover ( $n = 13$ ); nucleotide transport and metabolism ( $n = 11$ ); defense mechanisms ( $n = 5$ ); cell cycle control, cell division, and chromosome partitioning ( $n = 9$ ); and RNA processing and modification ( $n = 2$ ).

We identified a mutation on the *rpoB* gene (D516V) related to resistance to rifampin (11), mutations on the *katG* (S315T) and *kasA* genes (G269S) related to resistance to isoniazid (12, 13), and a mutation on the *pncA* gene (Q10R) related to pyrazinamide resistance (14). Also, we found mutations on the *gyrA* gene (E21Q, S95T, G247S, G668D) and a mutation on the *embB* gene (Y319S) that do not confer resistance (15–17).

Comparison of the KZN 1435 and INS-MDR strains demonstrated that, despite the differences in geographic origin and the high incidence of TB in South Africa (1), the strains have similar proportions of SNPs, showing high degrees of

conservation of genome structure. These results suggest that the outbreak of drug-resistant tuberculosis in Lima may be a clonal expansion of the same strain; however, more genomic information is required.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number **JAQI00000000**. The version described in this paper is version JAQI01000000.

## ACKNOWLEDGMENT

---

This study was supported by the Peruvian National Institute of Health.

## FOOTNOTES

---

Address correspondence to H. Guio, [heinnerguido@gmail.com](mailto:heinnerguido@gmail.com).

**Citation** Galarza M, Tarazona D, Borda V, Agapito JC, Guio H. 2014. Evidence of clonal expansion in the genome of a multidrug-resistant *Mycobacterium tuberculosis* clinical isolate from Peru. *Genome Announc.* 2(1):e00089-14. doi:10.1128/genomeA.00089-14.

**Received** 23 January 2014.

**Accepted** 3 February 2014.

**Published** 27 February 2014.

Copyright © 2014 Galarza et al.

This is an open-access article distributed under the terms of the **Creative Commons Attribution 3.0 Unported license**.

## REFERENCES

---

1. World Health Organization. 2013. Global tuberculosis report 2013. World Health Organization, Geneva, Switzerland.
2. ESN PCT. 2013. Informe de la Estrategia Sanitaria Nacional de Prevención y Control de la Tuberculosis sobre Situación y Control de TB-MDR y TB-XDR en Perú. <http://190.40.40.245/newtb/Archivos/RecursoInformacion/20140211114915.pdf>.
3. Abubakar I, Zignol M, Falzon D, Raviglione M, Ditiu L, Masham S, Adetifa I, Ford N, Cox H, Lawn SD, Marais BJ, McHugh TD, Mwaba P, Bates M, Lipman M, Zijenah L, Logan S, McNerney R, Zumla A, Sarda K, Nahid P, Hoelscher M, Pletschette M, Memish ZA, Kim P, Hafner R, Cole S, Migliori GB, Maeurer M, Schito M, Zumla A. 2013. Drug-resistant tuberculosis: time for visionary political leadership. *Lancet Infect. Dis.* 13:529–539. 10.1016/S1473-3099(13)70030-6. CrossRefMedlineGoogle Scholar
4. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 44:4498–4510. 10.1128/JCM.01392-06. Abstract/FREE Full Text
5. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, Bobadilla del Valle M, Fyfe J, García-García L, Rastogi N, Sola C, Zozio T, Guerrero MI, León CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendón A, Sifuentes-Osornio J, Ponce de León A, Cave MD, Fleischmann R, Whittam TS, Alland D. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 188:759–772. 10.1128/JB.188.2.759-772.2006. Abstract/FREE Full Text
6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. 10.1093/bioinformatics/btp324. Abstract/FREE Full Text
7. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. 10.1186/1471-2105-9-386. CrossRefMedlineGoogle Scholar
8. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. 10.1186/1471-2105-4-41. CrossRefMedlineGoogle Scholar

9. Ioerger TR, Koo S, No E-G, Chen X, Larsen MH, Jacobs WR Jr, Pillay M, Sturm AW, Sacchettini JC. 2009. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* 4:e7778. 10.1371/journal.pone.0007778. CrossRefMedlineGoogle Scholar
10. Song J, Xu Y, White S, Miller KW, Wolinsky M. 2005. SNPsFinder—a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. *Bioinformatics* 21:2083–2084. 10.1093/bioinformatics/bti176. Abstract/FREE Full Text
11. Lipin MY, Stepanshina VN, Shemyakin IG, Shinnick TM. 2007. Association of specific mutations in *katG*, *rpoB*, *rpsL* and *rrs* genes with spoligotypes of multidrug-resistant *Mycobacterium tuberculosis* isolates in Russia. *Clin. Microbiol. Infect.* 13:620–626. 10.1111/j.1469-0691.2007.01711.x. CrossRefMedlineGoogle Scholar
12. Campbell PJ, Morlock GP, Sikes RD, Dalton TL, Metchock B, Starks AM, Hooks DP, Cowan LS, Plikaytis BB, Posey JE. 2011. Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 55:2032–2041. 10.1128/AAC.01550-10. Abstract/FREE Full Text
13. Lee AS, Lim IH, Tang LL, Telenti A, Wong SY. 1999. Contribution of *kasA* analysis to detection of isoniazid-resistant *Mycobacterium tuberculosis* in Singapore. *Antimicrob. Agents Chemother.* 43:2087–2089. Abstract/FREE Full Text
14. Cheng SJ, Thibert L, Sanchez T, Heifets L, Zhang Y. 2000. *pncA* mutations as a major mechanism of pyrazinamide resistance in *Mycobacterium tuberculosis*: spread of a monoresistant strain in Quebec, Canada. *Antimicrob. Agents Chemother.* 44:528–532. 10.1128/AAC.44.3.528-532.2000. Abstract/FREE Full Text
15. Plinke C, Rüsche-Gerdes S, Niemann S. 2006. Significance of mutations in *embB* codon 306 for prediction of ethambutol resistance in clinical *Mycobacterium tuberculosis* isolates. *Antimicrob. Agents Chemother.* 50:1900–1902. 10.1128/AAC.50.5.1900-1902.2006. Abstract/FREE Full Text
16. Lau RW, Ho PL, Kao RY, Yew WW, Lau TC, Cheng VC, Yuen KY, Tsui SK, Chen X, Yam WC. 2011. Molecular characterization of fluoroquinolone resistance in *Mycobacterium tuberculosis*: functional analysis of *gyrA* mutation at position 74. *Antimicrob. Agents Chemother.* 55:608–614. 10.1128/AAC.00920-10. Abstract/FREE Full Text
17. Malik S, Willby M, Sikes D, Tsodikov OV, Posey JE. 2012. New insights into fluoroquinolone resistance in *Mycobacterium tuberculosis*: functional genetic analysis of *gyrA* and *gyrB* mutations. *PLoS One* 7:e39754. 10.1371/journal.pone.0039754. CrossRefMedlineGoogle Scholar



# Genome Analysis of 17 Extensively Drug-Resistant Strains Reveals New Potential Mutations for Resistance

H. Guio<sup>a,c</sup>, D. Tarazona<sup>a</sup>, M. Galarza<sup>a</sup>, V. Borda<sup>a</sup>, R. Curitomay<sup>b</sup>

Author Affiliations

<sup>a</sup> Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud (INS), Lima, Peru

<sup>b</sup> Departamento de Neumología, Hospital Nacional Hipólito Unanue, Lima, Peru

<sup>c</sup> Asociación Latinoamericana de Biotecnología (NGO-ALBIOTEC), Lima, Peru

## ABSTRACT

We report the whole-genome sequence of an extensively drug-resistant (XDR) tuberculosis (TB) strain of Latin American–Mediterranean (LAM) lineage. This strain is phenotypically resistant to aminoglycosides, but carries no related mutations in *rrs*, *tlyA*, and *eis*. Through genome analysis comparison with 16 XDR strains, we found 218 non-synonymous single nucleotide polymorphisms (SNPs) shared that could confer resistance.

## GENOME ANNOUNCEMENT

For Peru, the incidence rate for tuberculosis (TB) in 2012 was 95/100,000 population, from which 0.3% were new cases of extensively drug-resistant (XDR) TB. 94% of these cases were from Lima (1). XDR-TB strains are resistant to isoniazid (INH), rifampicin (RIF), fluoroquinolones (FLQ), and at least one of the three injectable anti-TB drugs: amikacin (AMK), capreomycin (CAP), or kanamycin (KAN) (2). In 1999, Peru reported one case with this resistant pattern. However, the concept of XDR-TB appears in 2006 with the first outbreak in KwaZulu-Natal (KZN), South Africa (3).

In our quest to identify new possible mutations that can confer resistance in XDR-TB strains, we analyzed the XDR strain, INS-XDR, isolated from a Peruvian patient with active tuberculosis. We determined the strain's lineage to be Latin American–Mediterranean (LAM) using 24 mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) loci (4) and SpolPred (5). The genomic DNA was sequenced on an Illumina HiSeq 2000 with 1,140× coverage. Sequence reads for the isolates were mapped against the corrected reference genome *M. tuberculosis* strain H37Rv. The average coverage was 82 for INS-XDR (49,793,402 mapped reads) and assembly with BWA v0.5.9-r16 (6), using H37Rv genome (AL123456.3) (7) as a reference. For annotation process, we used Rapid Annotations using Subsystem Technology (8) and the Prokaryotic Genome Annotation Pipeline. The sequencing data generated has 49,793,402 paired-ends reads. The assembling process resulted in 20 contigs that represents 99.98% of the H37Rv genome. The INS-XDR has a total of 4,391,020 pb with an average G+C content of 65.4%. We used Sanger sequencing to confirm specific mutations. Phenotypically, this strain was resistant to RIF, INH, FLQ, KAN, and CAP. However, sequence analysis of *rrs* and the promoter *eis* do not show mutations for aminoglycosides resistance (9).

To analyze possible mutations causing resistance to CAP and KAN, a polymorphism study was carried out by comparing INS-XDR against KZN605 (a XDR-TB strain of LAM lineage) and 16 reported TB-XDR genomes from China, FJ05194, GuangZ0019 (10), XDR1219, XDR1221 (11), GuangZ0008, 293, FJ07070, GuangZ0017, GuangZ0026 (12); Peru, Peru\_04\_R0292, Peru\_03\_R0268 (13); Russia, CTRI-4 (14), Russia\_03-R1082 (13); Malaysia, UM1072388579 (15), and South Africa, KZN\_R506 (16) and KZN\_605. Single nucleotide polymorphisms (SNPs) were identified using SNPsFinder (17) and SNP tree (18), and were classified in the Clusters of Orthologous Groups (COG) database (19). These analyses showed evidence of 218 non-synonymous SNPs mainly in four COGs: (1) replication, recombination and repair, L, *n*=14, (2) energy production and conversion, C, *n*=14, (3) amino acid transport and metabolism, E, *n*=12, and (4) secondary metabolites biosynthesis, transport, and catabolism, Q, *n*=11. These 218 non-synonymous SNPs belong to 125 genes. Given that CAP and KAN act by inhibiting protein synthesis and causing production of abnormal proteins, we hypothesize that the 14 non-synonymous SNPs belonging to the C category would be the principal SNPs responsible for CAP and KAN resistance. This provides evidence that it may be worthwhile to explore these new SNPs for diagnosis of XDR-TB.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession **JANH00000000**. The version described in this paper is version JANH01000000.

## ACKNOWLEDGMENTS

---

We thank Kelly Levano and Harrison Montejo.

This work was supported and approved by the research and ethical committee from the Peruvian National Institute of Health.

## FOOTNOTES

---

Address correspondence to H. Guio, [hguio@ins.gob.pe](mailto:hguio@ins.gob.pe).

**Citation** Guio H, Tarazona D, Galarza M, Borda V, Curitomay R. 2014. Genome analysis of 17 extensively drug-resistant strains reveals new potential mutations for resistance. *Genome Announc.* 2(4):e00759-14. doi:10.1128/genomeA.00759-14.

**Received** 7 July 2014.

**Accepted** 16 July 2014.

**Published** 31 July 2014.

Copyright © 2014 Guio et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

## REFERENCES

---

1. ESNPyCTB. 2012. Informe de la Estrategia Sanitaria Nacional de Prevención y Control de la Tuberculosis sobre Situación y Control de TB-MDR y TB-XDR en Perú. Estrategia Sanitaria Nacional de Prevención y Control de la Tuberculosis.
2. Jassal M, Bishai WR. 2009. Extensively drug-resistant tuberculosis. *Lancet Infect. Dis.* 9:19–30. 10.1016/S1473-3099(08)70260-3. CrossRefMedlineGoogle Scholar
3. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, Zeller K, Andrews J, Friedland G. 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* 368:1575–1580. 10.1016/S0140-6736(06)69573-1. CrossRefMedlineGoogle Scholar
4. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Loch C, van Soolingen D. 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 44:4498–4510. 10.1128/JCM.01392-06. Abstract/FREE Full Text
5. Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNERNEY R, Martin N, Clark TG. 2012. SpolPred: Rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. *Bioinformatics* 28:2991–2993. 10.1093/bioinformatics/bts544. Abstract/FREE Full Text
6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. 10.1093/bioinformatics/btp324. Abstract/FREE Full Text
7. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45:1176–1182. 10.1038/ng.2744. CrossRefMedlineGoogle Scholar
8. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42:D206–D214. 10.1093/nar/gkt1226. Abstract/FREE Full Text
9. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. 2009. Tuberculosis drug resistance mutation database. *PLOS Med.* 6:e2. 10.1371/journal.pmed.1000002. CrossRefMedlineGoogle Scholar
10. Lin N, Liu Z, Zhou J, Wang S, Fleming J. 22 August 2013. Draft genome sequences of two super-XDR isolates of *Mycobacterium tuberculosis* from China. *FEMS Microbiol. Lett.* 10.1111/1574-6968.12238. CrossRefGoogle Scholar
11. Wu W, Zheng H, Zhang L, Wen Z, Zhang S, Pei H, Yu G, Zhu Y, Cui Z, Hu Z, Wang H, Li Y. 2013. A genome-wide analysis of multidrug-resistant and extensively drug-resistant strains of *Mycobacterium tuberculosis* Beijing genotype. *Mol. Genet. Genomics* 288:425–436. 10.1007/s00438-013-0758-4. CrossRefGoogle Scholar

12. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang X-E, Bi L. 2013. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat. Genet.* 45:1255–1260. 10.1038/ng.2735. CrossRefMedlineGoogle Scholar
13. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PK, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* 45:1183–1189. 10.1038/ng.2747. CrossRefMedlineGoogle Scholar
14. Ilina EN, Shitikov EA, Ikryannikova LN, Alekseev DG, Kamashev DE, Malakhova MV, Parfenova TV, Afanas'ev MV, Ischenko DS, Bazaleev NA, Smirnova TG, Larionova EE, Chernousova LN, Beletsky AV, Mardanov AV, Ravin NV, Skryabin KG, Govorun VM. 2013. Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia. *PLoS One* 8:e56577. 10.1371/journal.pone.0056577. CrossRefMedlineGoogle Scholar
15. Ng KP, Yew SM, Chan CL, Chong J, Tang SN, Soo-Hoo TS, Na SL, Hassan H, Ngeow YF, Hoh CC, Lee KW, Yee WY. 2013. Draft Genome Sequence of the first Isolate of Extensively drug-Resistant (XDR) *Mycobacterium tuberculosis* in Malaysia. *Genome Announc.* 1(1):e00056-12. 10.1128/genomeA.00056-12. Abstract/FREE Full Text
16. Iøerger TR, Koo S, No E-G, Chen X, Larsen MH, Jacobs WR Jr, Pillay M, Sturm AW, Sacchettini JC. 2009. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* 4:e7778. 10.1371/journal.pone.0007778. CrossRefMedlineGoogle Scholar
17. Song J, Xu Y, White S, Miller KW, Wolinsky M. 2005. SNPsFinder—a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. *Bioinformatics* 21:2083–2084. 10.1093/bioinformatics/bti176. Abstract/FREE Full Text
18. Leekitcharoenphon P, Kaas RS, Thomsen MC, Friis C, Rasmussen S, Aarestrup FM. 2012. snpTree—a web-server to identify and construct SNP trees from whole genome sequence data. *BMC Genomics* 13(Suppl 7):S6. 10.1186/1471-2164-13-S7-S6. CrossRefGoogle Scholar
19. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. 10.1186/1471-2105-4-41. CrossRefMedlineGoogle Scholar

Rev Peru Med Exp Salud Publica. 2016 Jun;33(2):256-63.

## Análisis genómico comparativo de cepas peruanas de *Mycobacterium tuberculosis*

Comparative genomic analysis of peruvian strains of *Mycobacterium tuberculosis*

David Tarazona<sup>1,a,c</sup>, Marco Galarza<sup>1,a</sup>, Kelly S. Levano<sup>1,b</sup>, Heinner Guio<sup>1,d</sup>

<sup>1</sup> Laboratorio de Referencia Nacional de Biología Molecular y Biotecnología, Centro Nacional de Salud Pública, Instituto Nacional de Salud, Lima, Perú.

<sup>a</sup> Biólogo molecular; <sup>b</sup> PhD en Bioquímica; <sup>c</sup> Genetista; <sup>d</sup> PhD en Ciencias Médicas

---

### RESUMEN

**Objetivos.** Analizar comparativamente tres secuencias genómicas de *Mycobacterium tuberculosis*(MTB): INS-SEN, cepa sensible; INS-MDR, cepa multidrogorresistente e INS-XDR, cepa extensamente resistente, procedentes de la Ciudad de Lima, Perú. **Materiales y métodos.** Se identificaron los polimorfismos de un solo nucleótido (SNPs) específicos en las cepas INS-SEN, INS-MDR y INS-XDR mediante el criterio de inclusión/exclusión. Se compararon los tres genomas de MTB y se construyó una filogenia molecular con 27 cepas de MTB de otros estudios, disponibles de la base de datos Genbank. Los SNPs específicos en cada genoma fueron organizados en clústers de grupos ortólogos (COGs). **Resultados.** El análisis de genomas permitió identificar un conjunto de SNPs asociados a determinantes de virulencia (familia de proteínas mce, policetidos, phiRv1, transposasas, metiltransferasas y relacionados a síntesis de vitaminas) principalmente. Se observa una estrecha relación entre la cepa INS-MDR y INS-XDR, con solo un 6,1% de SNPs diferentes, sin embargo, la cepa INS-SEN presenta un 50,2 y 50,3% de SNPs diferentes a las cepas MDR y XDR, respectivamente. La filogenia molecular agrupó a las cepas peruanas dentro del linaje LAM y cercanamente a las cepas F11 y KZN de Sudáfrica. **Conclusiones.** Se evidenció una alta similitud (99,9%) de la cepa INS-SEN con la cepa sudafricana F11, de gran alcance mundial, mientras los análisis de las cepas INS-MDR e INS-XDR demuestran una probable expansión de la familia KZN, cepa de Sudáfrica con alta virulencia y patogenicidad.

**Palabras clave:** Tuberculosis; Genómica; Farmacorresistencia bacteriana (fuente: DeCS BIREME).

---

### ABSTRACT

**Objectives.** To comparatively analyze three genomic sequences of *Mycobacterium tuberculosis*(MTB), including sensitive (INS-SEN), multi-drug-resistant (INS-MDR), and extremely drug-resistant (INS-XDR) strains, collected in Lima, Peru. **Materials and Methods.** Specific single nucleotide polymorphisms (SNPs) were identified in the INS SEN, INS-MDR, and INS-XDR strains according to the inclusion/exclusion criteria. The three MTB genomes were compared and a molecular phylogeny was constructed with 27

MTB strains from other studies available from the Genbank database. **Results.** The specific SNPs in each genome were organized in clusters of orthologous groups (COGs). The genomic analysis allowed for the identification of a set of SNPs associated mainly with virulence determinants (family of mce proteins, polyketides, phiRv1, transposase, and methyltransferases, and other related to vitamin synthesis). A close correlation between the INS-MDR and INS-XDR strains was observed, with only a 6.1% difference in SNPs; however, the INS-SEN strain had 50.2% and 50.3% different SNPs from the MDR and XDR strains, respectively. The molecular phylogeny grouped the Peruvian strains within the LAM lineage and closely to the F11 and KZN strains from South Africa. **Conclusions.** High similarity (99.9%) was noted between the INS-SEN strain and the F11 South African strain with broad global scope, while the analysis of the INS-MDR and INS-XDR strains showed a likely expansion of the KZN family, a South African strain with high virulence and pathogenicity.

**Key words:** Tuberculosis; Genomic; Drug Resistant, Bacterial (source: MeSH NLM).

---

## INTRODUCCIÓN

Según la Organización Mundial de la Salud (OMS), durante el 2013 se han reportado 6,1 millones de casos de tuberculosis (TB). De estos, el 3,5% correspondieron a casos nuevos de tuberculosis multidrogorresistente (TB-MDR) y el 9% de estos casos han constituido casos nuevos de tuberculosis extensamente resistente (TB-XDR). En Perú, la alta carga de TB y el incremento de cepas multidrogorresistentes (MDR) es uno de los mayores problemas en salud pública. En el año 2013, la tasa de incidencia de TB en Perú fue de 124/100 000 habitantes, de los cuales un 3,9% (850) fueron casos nuevos de TB-MDR <sup>(1)</sup>.

Si bien los casos nuevos de TB han disminuido en los últimos años, los casos de TB-MDR se han incrementado en los últimos 3 años. Durante el año 2014, Perú representó el 41,3% de todos los casos TB-MDR de la región de las Américas <sup>(2)</sup>. Lima concentra el 80 y 92% de casos TB-MDR y TB-XDR respectivamente <sup>(3)</sup>, donde un factor importante es el transporte público peruano, como contribuyente de factor de riesgo para TB y su diseminación <sup>(4)</sup>.

La TB-MDR está definida como la resistencia a rifampicina (RIF) e isoniácida (INH), dos de las drogas más potentes usadas en el tratamiento antituberculoso.

La TB-XDR definida como la resistencia a RIF, INH, fluoroquinolonas y, al menos, uno de los antibióticos inyectables (amikacina, kanamicina o capreomicina).

La RIF es un antibiótico bactericida que inhibe la enzima ARN polimerasa de la micobacteria. La región genética relacionada a resistencia a RIF está localizada en una región de 81 pb del gen *rpoB* de *Mycobacterium tuberculosis* (MTB). La INH es una prodroga que es activada por la enzima *katG* (catalasa peroxidasa) de MTB, y actúa inhibiendo la síntesis de ácido micólico. La resistencia a INH es frecuentemente observada en el codón 315 del gen *katG*, además de otras importantes regiones de mutación que pueden ocurrir en los genes *ahpC*, *kasA*, *ndh* y la región promotora *inhA* <sup>(5)</sup>.

Existe una alta diversidad genética de cepas de MTB a nivel mundial que causan la enfermedad. Se han descrito como prevalentes en el mundo los siguientes genotipos: Haarlem (H), Latin América-Mediterráneo (LAM), T, Este de África-India (EAI), Central-Asia (CAS), X y Beijing. Cada nombre está atribuido al lugar de procedencia donde se identificó inicialmente. Asimismo, los recientes estudios en análisis de genomas están

identificando nuevas regiones en MTB como posibles sitios polimórficos en relación a resistencia a drogas antituberculosas <sup>(6)</sup>.

La evolución de cepas multidrogorresistentes (MDR) y extensamente resistente (XDR) tiene una estrecha relación a cepas provenientes de Sudáfrica, como las familias F11 y KZN. La familia F11 data de los años 1992- 1998, provenientes del oeste de Cape en Sudáfrica y que representó el 21,4% de todos los casos de TB en esos años. Presentan como características principales: la presencia de bandas patrones de la región *IS6110* que varía de 11 a 19 bandas, ausencia de regiones espaciadoras 9 a 11; 21 a 24 y 33 a 36, así como la presencia de polimorfismo C491T en el gen *rrs* <sup>(7)</sup>. Por otro lado, la familia KZN proviene de la región de KwaZulu-Natal en Sudáfrica, y data de un brote de los años 2005-2007. Es una cepa altamente virulenta que causó 52 muertes de 53 pacientes con diagnóstico TB-XDR y HIV <sup>(8)</sup>

Existe evidencia que poblaciones nativas peruanas han sido afectados por TB <sup>(9)</sup>. Sin embargo, el genotipo de MTB que los afecta aún permanece desconocido. Se tiene registrado que en Perú existen familias genéticas Beijing desde la llegada de migrantes asiáticos en el decenio de 1900 <sup>(10)</sup>, y que recientemente fue reportado genotipos similares en España e Italia <sup>(11)</sup>. El primer estudio de la diversidad genética de MTB en el Perú de aislados sensible y drogorresistentes, indica que los genotipos predominantes son: LAM (23,8%), T (23,8%), Haarlem (22,3%) y Beijing (9,3%) <sup>(12)</sup>. Sheen *et al.* reportan como prevalentes a los linajes Haarlem (28,7%), LAM (28,3%) y T (20,3%) en pacientes con TB y VIH <sup>(13)</sup>. Investigadores del Instituto Nacional de Salud (INS) de Perú, identificaron que pacientes con TB-XDR presentaron los linajes Haarlem (43,6%), T (27,7%), LAM (16,2%) y Beijing (9,5%) <sup>(3)</sup>. Estos estudios permiten una primera aproximación, de los genotipos LAM, Haarlem y T en Perú. Grandjean *et al.* (2015), en un estudio con 2139 aislados, encontraron que la familia LAM puede representar al menos el 50% de todos los casos de resistencia a droga en la región peruana <sup>(14)</sup>.

Durante el 2014, se realizó el secuenciamiento completo de tres genomas de MTB (sensible, MDR y XDR) aislados en Perú, de linaje LAM. El presente estudio tiene como objetivo realizar un análisis comparativo de los tres genomas frente a otros genomas referenciales, y poder establecer su relación genética entre ellos.

## **MATERIALES Y MÉTODOS**

### *CEPAS, SENSIBILIDAD A DROGAS Y PURIFICACIÓN DE ADN*

Los aislados fueron cultivados en medio solido Löwenstein-Jensen y la susceptibilidad a drogas de primera y segunda línea fue por el método estándar Agar Proporciones en Placa (APP), en el Laboratorio de Referencia Nacional de Micobacterias del Instituto Nacional de Salud. Se realizó la extracción de ADN usando el kit PureLink® Genomic DNA (Invitrogen), y la cuantificación del ADN se hizo en el equipo NanoDrop 8000 (Thermo Scientific) del Laboratorio de Referencia Nacional de Biotecnología y Biología Molecular del INS. En la Tabla 1 se muestran las características básicas de los aislados de MTB sensible, multidrogorresistente y extremadamente resistente, seleccionados para el estudio. Mediante la metodología MIRU-VNTR (*Mycobacterial Interspersed Repetitive Unit-Variable Number Of Tandem Repeat*) se identificó el linaje de MTB.

**Tabla 1.** Descripción de las cepas peruanas de *Mycobacterium tuberculosis* que forman parte de este estudio

Cepa	Fenotipo	Genotipo (SNPs)	MIRU-VNTR*
INS-SEN	RIF <sup>s</sup> , INH <sup>s</sup> , EMB <sup>s</sup> , PZA <sup>s</sup> , ETH <sup>s</sup> , AMI <sup>s</sup> , CAP <sup>s</sup> , OFL <sup>s</sup>	-	46[14]337645758468357578756
INS-MDR	RIF <sup>r</sup> , INH <sup>r</sup> , EMB <sup>r</sup> , PZA <sup>r</sup> , ETH <sup>r</sup> , AMI <sup>s</sup> , CAP <sup>s</sup> , OFL <sup>s</sup>	<i>rpoB</i> (D516V), <i>katG</i> (S315T), <i>kasA</i> (G269S), <i>thyA</i> (T202A), <i>pncA</i> (Q10R)	13[11]224342225237264433522
INS-XDR	RIF <sup>r</sup> , INH <sup>r</sup> , EMB <sup>r</sup> , PZA <sup>r</sup> , ETH <sup>r</sup> , KAN <sup>r</sup> , CAP <sup>r</sup> , OFL <sup>r</sup>	<i>rpoB</i> (D516V), <i>katG</i> (S315T), <i>kasA</i> (G269S), <i>gyrA</i> (D94G), <i>thyA</i> (T202A), <i>pncA</i> (Q10R), <i>embB</i> (Y319S)	23[11]225353335238264523533

RIF: rifampicina INH: isoniazida, EMB: etambutol, PZA: pirazinamida, ETH: etionamida, KAN: kanamicina, CAP: capreomicina, OFL: ofloxacina.

\*24-VNTR: MIRU02, Mtub04, ETRC, MIRU04, MIRU40, MIRU10, MIRU16, Mtub21, MIRU20, QUB11b, ETRA, Mtub29, Mtub30, ETRB, MIRU23, MIRU24, MIRU26, MIRU27, Mtub34, MIRU31, Mtub39, QUB26, QUB4156 y MIRU39.

### DATOS GENÓMICOS DE CEPAS DE *Mycobacterium tuberculosis*

Se construyeron librerías genómicas de extremos pareados, las cuales fueron secuenciadas en el equipo HiSeq 2000, Illumina (Macrogen, Korea)<sup>(15-17)</sup>, siguiendo los protocolos del fabricante. El ensamblaje de la secuenciación de lecturas se llevó a cabo por el software BWA versión v 0.5.9-r16, usando el genoma referencial H37Rv.

### ANÁLISIS DE SECUENCIAS Y FILOGENIA MOLECULAR

Determinamos los SNPs estrechamente relacionados a MTB H37Rv (AL123456.2) mediante los programas SamTools, SNPtree y scripts en Perl para INS-SEN<sup>(15)</sup>, INS-MDR<sup>(16)</sup> e INS-XDR<sup>(17)</sup>. La anotación de secuencias codificantes en INS-SEN, INS-MDR y INS-XDR fue realizada usando los programas RAST (Rapid Annotations using Subsystem Technology) y PGAAP (*Prokaryotic Genome Annotation Pipeline*). Adicionalmente, las proteínas se agruparon en clúster de grupos ortólogos (COG) mediante BLAST y se organizaron los SNP en COG.

Se realizó un análisis comparativo por criterio de inclusión/exclusión usando un total de 785, 805 y 814 SNPs de INS-SEN, INS-MDR y INS-XDR, respectivamente, empleando las bases de datos, y elaboradas con *pipelines en Perl* - Linux. Adicionalmente, se elaboró una filogenia por *Máximo Likelihood* de los genomas de MTB listados en la [Tabla 2](#), para ello se utilizó el programa FastTree y se utilizó el programa de graficas FigTree.



**Tabla 2.** Descripción de las 30 cepas de *Mycobacterium tuberculosis* consideradas en este estudio

N.º	Genoma MTB	Número de accesoión	Tipo	Familia	Pais	Institución Responsable
1	H37Rv	AL123456.2	Sensible	-	USA	Wellcome Trust Sanger Institute
2	CTRI-2	NC_017524	Sensible	LAM	Rusia	Research Institute of Physical-Chemical Medicine
3	KZN-4207	NC_016768	Sensible	LAM	Sudáfrica	The University of Texas
4	KZN-1435	NC_012943	MDR	LAM	Sudáfrica	Broad Institute of MIT and Harvard
5	KZN-V2475	ACVT00000000	MDR	LAM	Sudáfrica	The University of Texas
6	KZN-R506	ACVU00000000	XDR	LAM	Sudáfrica	The University of Texas
7	KZN-605	ABGN00000000	XDR	LAM	Sudáfrica	Broad Institute of MIT and Harvard
8	W-148	ACSX01000000	MDR	Beijing	Rusia	Broad Institute of MIT and Harvard
9	X122	CM001044	XDR	Beijing	Sudáfrica	The University of Texas
10	CCDC 5079	NC_021251	Sensible	Beijing	China	Beijing Genomics Institute
11	210	ADAB01000000	Sensible	Beijing	USA	The Institute for Genomic Research
12	Guang Z0008	SRA065095	XDR	Beijing	China	Institute of Biophysics, Chinese Academy of Sciences
13	Guang Z0017	SRA065095	XDR	Beijing	China	Institute of Biophysics, Chinese Academy of Sciences
14	02_1987	ABLM00000000	ND	Beijing	USA	Broad Institute of MIT and Harvard
15	OSDD 518	AHHZ01000000	ND	CAS	India	CSIR Institute of Genomics and Integrative Biology
16	CAS-NITR 204	NC_021193	ND	CAS	India	National Institute for Research in Tuberculosis
17	OSDD 504	AHHY01000000	ND	CAS	India	CSIR Institute of Genomics and Integrative Biology
18	OSDD 071	AHHX01000018	ND	CAS	India	CSIR Institute of Genomics and Integrative Biology
19	EAI-OSDD 271	AQQC01000000	ND	EIA	India	CSIR Institute of Genomics and Integrative Biology
20	EAI5-NITR 206	CP005387	Sensible	EIA	India	CSIR Institute of Genomics and Integrative Biology
21	CDC 1551	NC_002755	Sensible	X	USA	The Institute for Genomic Research
22	Haarlem-NITR 202	CP004886	Sensible	Haarlem	India	National Institute for Research in Tuberculosis
23	Strain-C	AAKR01000000	Sensible	x	USA	Broad Institute of MIT and Harvard
24	Erdman ATCC 35801	AP012340	ND	Haarlem	USA	National Center for Global Health and Medicine
25	OSDD 493	AVQJ01000000	sensible	Ural	India	CSIR Institute of Genomics and Integrative Biology
26	F11	NC_009565	ND	LAM	Sudáfrica	Broad Institute of MIT and Harvard
27	HM	AZHK01000001	ND		Uruguay	Instituto Pasteur de Montevideo
28	INS-SEN	JAQH01000000	Sensible	LAM	Perú	Instituto Nacional de Salud
29	INS-MDR	JAQI01000000	MDR	LAM	Perú	Instituto Nacional de Salud
30	INS-XDR	JANH01000000	XDR	LAM	Perú	Instituto Nacional de Salud

## RESULTADOS

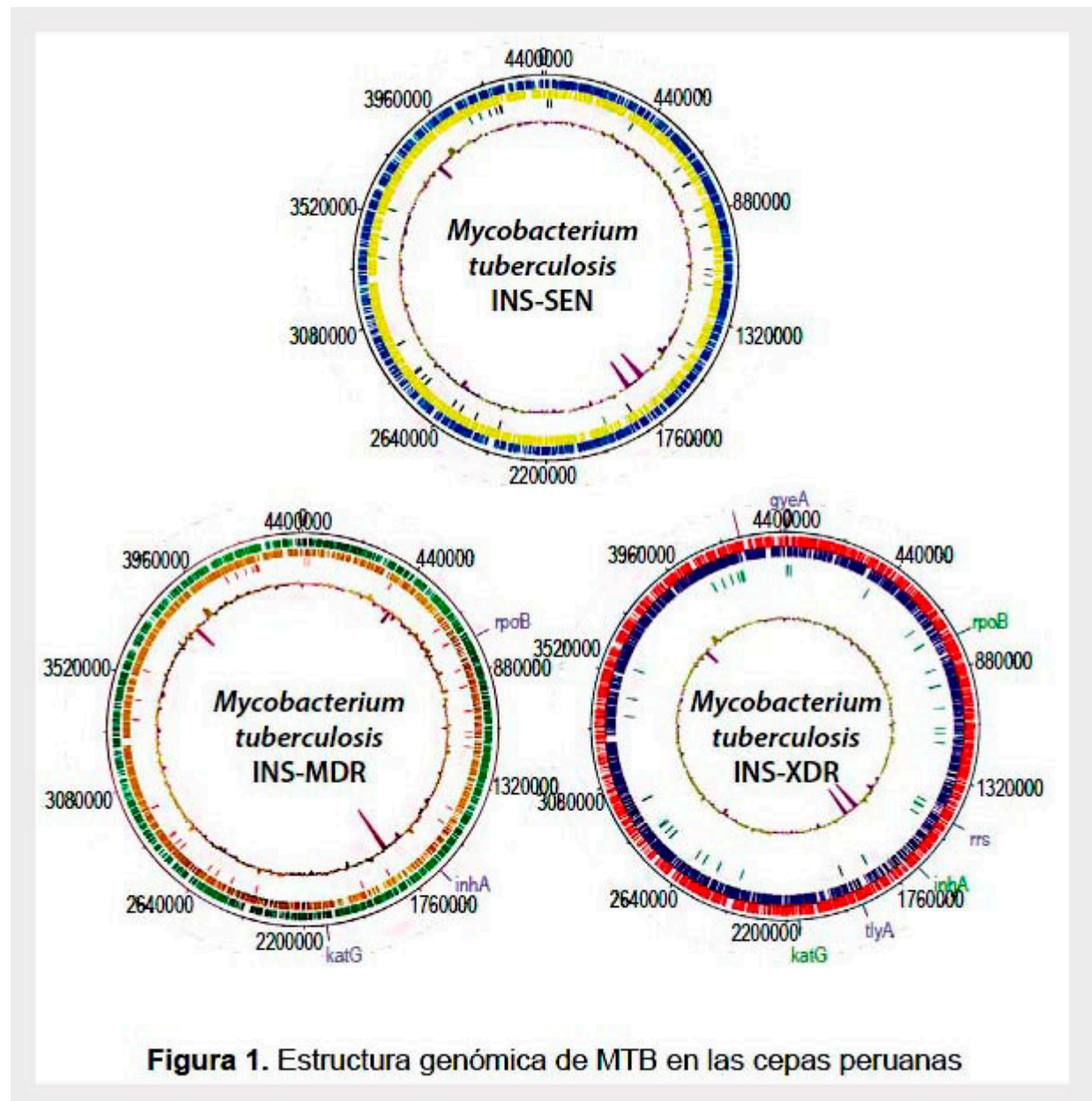
### ESTRUCTURA DE GENOMAS DE MTB EN AISLADOS PERUANOS

INS-SEN (NCBI: JAQH01000000) tuvo 61 422 158 lecturas de secuencias con 1406x de cobertura genómica logrando un tamaño de 4 383 671 pb y un contenido GC de 65,6%. INS-MDR (NCBI: JAQI01000000) de 58 157 302 lecturas de secuencias con 1331x de cobertura logrando un tamaño de 4383 671 pb y un contenido GC de 65,6%. INS-XDR (NCBI: JANH01000000) de 49 793 402 lecturas de secuencias con 1140x de cobertura logrando un tamaño de 4 391 020 pb y un contenido GC de 65,4%.

En nuestro análisis, encontramos que los tres genomas INS-SEN, INS-MDR y INS-XDR tienen 4389 secuencias codificantes, 45 tRNA, 03 tRNA (Figura 1). El estudio comparativo de cada genoma frente a H37Rv encontró que INS-SEN, INS-MDR y INS-XDR tienen 499, 805 y 815 SNPs, respectivamente. El porcentaje de SNPs localizados en regiones

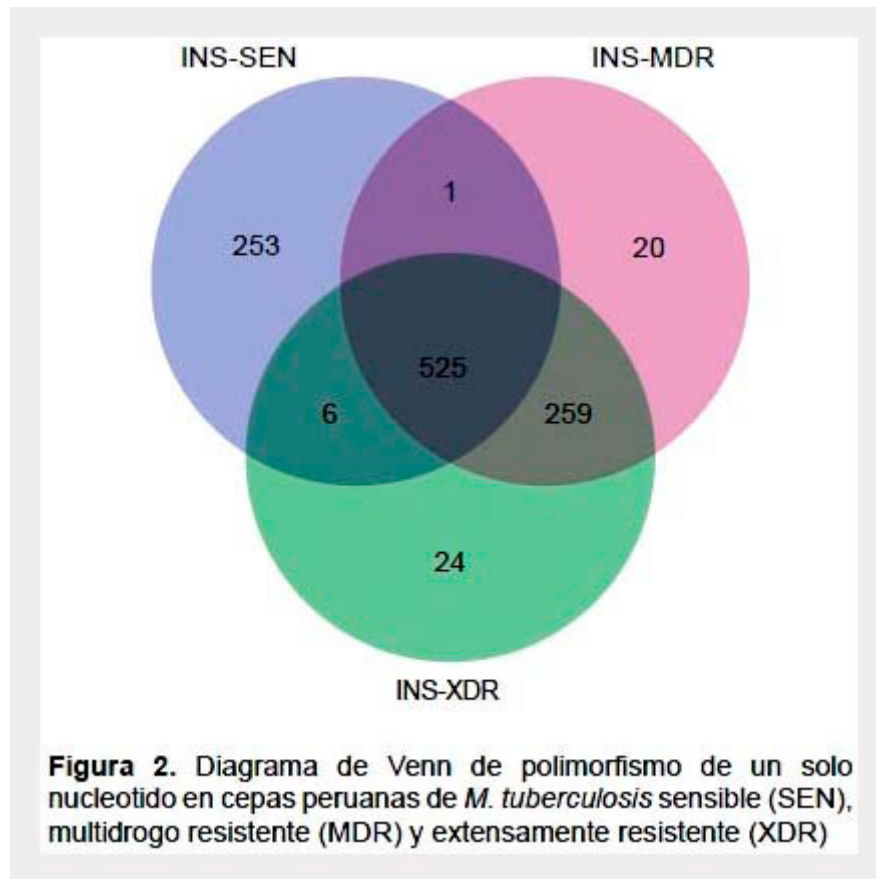


codificantes fueron, para INS-SEN (88,2%), INS-MDR (87,3%) y INS-XDR (87,3%) aproximadamente.



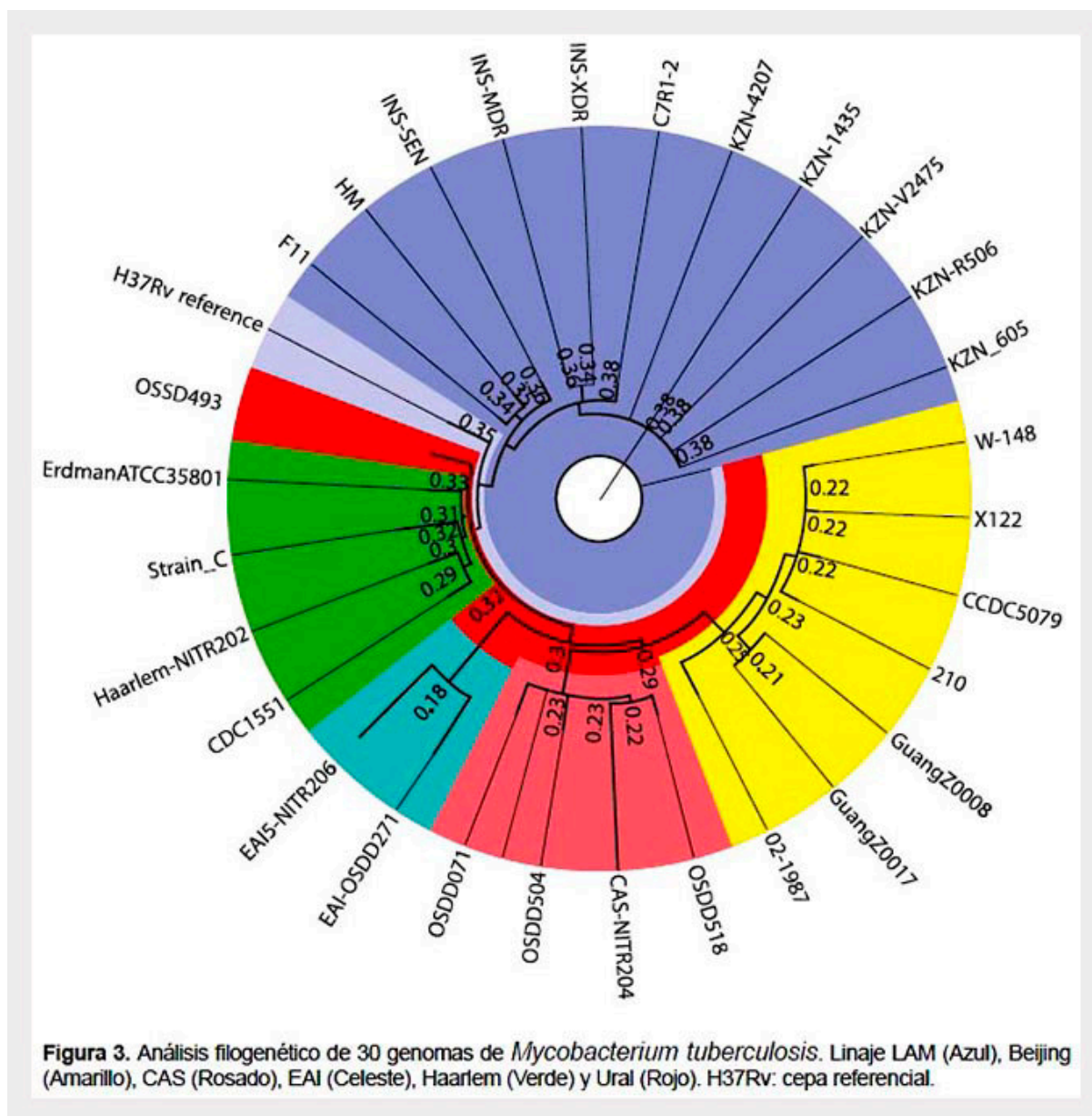
#### COMPARACIÓN DE SECUENCIAS DE MTB EN AISLADOS PERUANOS

La comparación de genomas de MTB de familia LAM por el criterio de inclusión/exclusión determinó que las cepas INS-SEN, INS-MDR y INS-XDR tienen SNPs específicos y compartidos (Figura 2). El número de SNPs propios fueron 253 para INS-SEN, 20 para INS-MDR y 24 para INS-XDR. Los SNPs compartidos para los tres genomas es de 48,3% (525/1088). El número de SNPs compartidos entre INS-SEN y INS-MDR es de 26,2%; INS-SEN y INS-XDR 26,4%.



#### FILOGENIA MOLECULAR DE CEPAS DE MTB

La filogenia molecular construida a partir de 30 genomas de MTB de familias representativas de linajes LAM, Beijing, CAS, Haarlem, EAI y Ural ([Figura 3](#)) ha mostrado ser eficaz para bacterias genéticamente monomórficos como MTB. La filogenia agrupó los aislados en linajes, sin embargo no se puede observar grupos bien diferenciados de cepas sensibles o drogo resistentes. Esta filogenia divide todos los aislados en diferentes grupos, estableciendo que INS-SEN se encuentra en el linaje LAM y más relacionado a HM y F11 (proveniente del brote en Cape en Sudáfrica).



En contraste los aislados, INS-MDR e INS-XDR también se encuentran en linaje LAM pero más relacionados a la familia KZN (proveniente del brote KwaZulu-Natal en Sudáfrica).

#### ANÁLISIS DE GRUPOS ORTÓLOGOS (COGS)

Los genomas en este estudio fueron: sensible, MDR, XDR y de linaje LAM. Se formó una lista de SNPs individual para cada genoma INS frente a H37Rv. Las proteínas codificadas que presentaron SNPs fueron asociadas a veinte clases funcionales basados en grupos de ortólogos (COG). Los SNPs se normalizaron con respecto al número de genes de un determinado COG en MTB y con el número total de SNPs identificados (Figura 4). Los grupos K: transcripción; J: traducción, biogénesis ribosomal; O: modificación postraduccional, chaperonas; V: mecanismo de defensa; U: secreción, tráfico intracelular y transporte vesicular, y A: procesamiento y modificación de ARN; no tuvieron un ajuste estructural sobre la asignación, lo que sugiere que los genes de estas categorías podrían estar bajo presión evolutiva. En todas las cepas peruanas, los números significativos de SNPs se encuentran en las proteínas Q: biosíntesis, transporte y catabolismo de

metabolismo secundario y L: replicación, recombinación y reparación. Dentro del grupo Q se encuentran los genes de *mce* y síntesis de policetidos. Mientras que en el grupo L encontramos genes de la ADN girasa, fago phiRv1, transposasas y ligasas, también metiltransferasas, más exclusivamente en las micobacterias drogorresistentes. El grupo H: metabolismos y transporte de coenzima, en la cepa sensible ha mostrado mayor valor significativo que las drogorresistentes (MDR y XDR), donde encontramos los genes que codifican enzimas para síntesis de riboflavina, cobalamina, glutamina, molibdeno, nicotinamida y magnesio quelatasa. Notamos también que el grupo G – metabolismo y transporte de carbohidratos es significativamente diferente entre la cepa sensibles y las drogorresistentes, dentro de este grupo de proteínas encontramos a las transcetolasas, *opcA*, transportador de azúcar, glucosidasas, permeasas de la familia MFS.

## DISCUSIÓN

En la actualidad, los avances en genómica comparativa permiten conocer en forma precisa, los polimorfismos relacionados a la fisiología, factores de virulencia y resistencia a drogas de los principales agentes infecciosos de enfermedades con alto impacto en salud pública. En este estudio se reporta por primera vez un análisis de tres genomas de MTB (sensible, MDR y XDR) que provienen de pacientes que solo han vivido en la ciudad de Lima, Perú, por lo que es razonable pensar que estos organismos representan una fracción de las cepas sensibles y drogorresistentes que circulan en Perú. En general, los aislados del MTB (INS-SEN, INS-MDR e INS-XDR), muestran un conjunto de SNP más asociados a los grupos funcionales Q y L, la familia Q está más representado por la familia de proteína *Mce*, un factor de virulencia importante en MTB. Tienen funciones en la inducción tráfico de sustrato en la fase de infección<sup>(18)</sup>, modificación del perfil lipídico y capacidad de supervivencia intracelular <sup>(19)</sup>. La disrupción funcional de *Mce1* causa hipervirulencia mientras las disrupciones en *Mce2* y *Mce3* confieren que los aislados sean atenuados <sup>(20)</sup>. En este estudio se evidenció que INS-SEN contiene una mutación adicional en Ser270Asn de *Mce2D* lo que podría relacionarse a su virulencia y atenuación. Adicionalmente, se encontró Val87Ala de *Mce1B* correspondiente solo a los genomas INS-MDR y INS-XDR que podrían ser los responsables de los cambios lipídicos en la membrana.

Mientras que en la familia L encontramos a los fagos PhiRv1, que tienen un sistema activo de integración/ escisión que les da facultades de cambiar de posiciones en el genoma de MTB <sup>(21)</sup>. El aislado INS-SEN ha mostrado más variabilidad en secuencias PhiRv1 que en los aislados drogorresistentes. Aunque no se ha reportado asociaciones de sensibilidad y drogorresistencia, existen reportes que indican que pueden variar con el estrés ambiental y mejorar el estado físico de MTB <sup>(22)</sup>, lo que podría ayudar a entender mejor su rol en los aislados sensibles. Otras secuencias como transposasas, están involucradas en eventos de microevolución, lo que confiere cambios funcionales en las adaptaciones de la cepa a un determinado hospedero <sup>(23)</sup>.

Nuestro reciente estudio del aislado sensible INS-SEN, indicó que los SNPs en la categoría COG tipo Q, I y L resultan ser más frecuentes, además de las secuencias codificantes de PPE y PE-PGRS. La importancia de las PPE radica en la asociación a la variación antigénica <sup>(24)</sup>, mecanismo usado por parte de organismos para alterar sus proteínas de superficie con el fin de evadir la respuesta inmune. Entre tanto, las PE-PGRS están asociadas a la variación antigénica y la evasión inmune <sup>(25)</sup>, estas proteínas ricas en secuencias GC repetidas, aumentan los eventos de recombinación y proporcionan variación génica. Las PPE y PE-PGRS tienen un importante rol en la virulencia de MTB <sup>(26)</sup>, lo que manifiesta sus patrones polimórficos de los aislados sensible y drogorresistente. INS-SEN posee una ventaja y es que probablemente tenga más virulencia frente a aislados MDR o XDR <sup>(27)</sup>. Se requieren más análisis para explorar los efectos de las mutaciones y su relación con la virulencia y la patogenicidad.

Los análisis de los genomas de MTB (MDR y XDR) contienen regiones de mutación diferentes a las relacionadas con resistencia a fármacos que resultan interesantes en

analizar, ya que pueden desempeñar un papel importante en la supervivencia, adaptación y su difusión. En este estudio, la cepa INS-MDR, de linaje LAM es resistente a RIF e INH, el cual presenta una alta similaridad con KZN1435, lo que evidencia una posible expansión de la familia KZN en nuestro país <sup>(16)</sup> e INS-XDR drogorresistente a RIF, INH, PZA, CAP, KAN y con alta similaridad a KZN 605 <sup>(17)</sup>. Estos hallazgos indican que INS-SEN y INS-XDR se encuentran estrechamente relacionados por evidencia filogenética que las ubican en el mismo lado con mínima distancia y los análisis comparativos de alta similaridad de SNPs entre ellas, con tan solo 51 nucleótidos diferentes, frente a un total de más de 4 millones de nucleótidos en cada genoma.

Los análisis en conjunto de cepas MDR y XDR determinan qué metiltransferasas se encuentran con más variabilidad, y esto debido probablemente a su rol en la reparación del ADN de MTB lo que provee una estabilidad genética, a pesar de la exposición del entorno hostil de macrófagos <sup>(28)</sup>, además de actuar como activador de drogas, por ejemplo a tiacetazona <sup>(29)</sup>. También el hecho de que genes involucrados en la síntesis de riboflavina, cobalamina, glutamina, molibdeno, nicotidamina y magnesio chelatasa se encuentren exclusivamente en las MDR y XDR, lleva a considerar que MTB y su cambio de *fitness* (eficacia biológica) presionan que existan modificaciones o selección dentro de las drogorresistentes para sintetizar moléculas clave en su funcionamiento normal o la modificación de alternativas en sus rutas metabólicas<sup>(30)</sup>. Mayores análisis son necesarios para determinar realmente la importancia de los SNPs en los aislados drogorresistentes, estos análisis nos acercarán más a lo que sucede en la dinámica del genoma de MTB drogorresistentes en los aislados peruanos.

En conclusión, los aislados peruanos identificados de MTB, de la familia LAM con sensibilidad a drogas y drogorresistentes (MDR y XDR) no se encuentran estrechamente relacionados. Además, que INS-SEN se encuentra relacionado a un brote F11 con 99,9% de similitud, mientras las cepas drogorresistentes INS-MDR y INS-XDR se encuentran asociadas al brote KZN de Sudáfrica. Identificamos SNPs presentes específicos de genomas sensible y para las secuencias genómicas drogorresistentes asociadas con virulencia y patogenicidad.

**Agradecimientos:** al Dr. Cesar Cabezas, exjefe del Instituto Nacional de Salud de Perú, quien durante su gestión apoyó para poder secuenciar las tres cepas peruanas. A la Dra. Lely Solari quien siendo Directora del Centro Nacional de Salud Pública nos apoyó en la realización de los análisis de las cepas. A Harrison Montejo del LBBM por su apoyo logístico.

**Contribuciones de autoría:** DT, MG, han participado en la concepción y diseño del artículo. Los procedimientos y resultados fueron realizados por DT, MG, los análisis y discusiones fueron realizados por DT, MG y KSL. La redacción del artículo estuvo a cargo DT, MG, KSL, HG. La revisión crítica la realizó HG. La versión final estuvo a cargo de HG.

**Fuentes de financiamiento:** este estudio fue financiado por el Centro Nacional de Salud Pública del Instituto Nacional de Salud de Perú.

**Conflictos de interés:** los autores declaran no tener conflictos de interés.

## REFERENCIAS BIBLIOGRÁFICAS

1. World Health Organization. Global Tuberculosis Report 2014. Geneva: WHO; 2014.

2. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384(9947):1005-70. doi: 10.1016/S0140-6736(14)60844-8.
3. Caceres O, Rastogi N, Bartra C, Couvin D, Galarza M, Asencios L, et al. Characterization of the genetic diversity of extensively-drug resistant *Mycobacterium tuberculosis* clinical isolates from pulmonary tuberculosis patients in Peru. *PloS One*. 2014;9(12):e112789. doi: 10.1371/journal.pone.0112789.
4. Garaycochea O, Ticona E. Rutas de transporte público y situación de la tuberculosis en Lima, Perú. *Rev Peru Med Exp Salud Publica*. 2015;32(1):93-7.
5. Somoskovi A, Parsons LM, Salfinger M. The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in *Mycobacterium tuberculosis*. *Respir Res*. 2001;2(3):164-8.
6. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet*. 2013;45(7):784-90. doi: 10.1038/ng.2656.
7. Victor TC, de Haas PE, Jordaan AM, van der Spuy GD, Richardson M, Van Soolingen D, et al. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a western cape F11 genotype. *J Clin Microbiol*. 2004;42(2):769-72.
8. Ioerger TR, Koo S, No EG, Chen X, Larsen MH, Jacobs WR Jr., et al. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One*. 2009;4(11):e7778. doi: 10.1371/journal.pone.0007778.
9. Culqui DR, Trujillo OV, Cueva N, Aylas R, Salaverry O, Bonilla C. Tuberculosis en la población indígena del Perú 2008. *Rev Peru Med Exp Salud Publica*. 2010;27(1):8-15.
10. Iwamoto T, Grandjean L, Arikawa K, Nakanishi N, Caviedes L, Coronel J, et al. Genetic diversity and transmission characteristics of Beijing family strains of *Mycobacterium tuberculosis* in Peru. *PloS One*. 2012;7(11):e49651. doi: 10.1371/journal.pone.0049651.
11. Garcia de Viedma D, Chaves F, Inigo J. New route of importation of *Mycobacterium tuberculosis* Beijing genotype. *Emerg Infect Dis*. 2006;12(1):169-70.
12. Taype CA, Agapito JC, Accinelli RA, Espinoza JR, Godreuil S, Goodman SJ, et al. Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Peru. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. *Infect Genet Evol*. 2012;12(3):577-85. doi: 10.1016/j.meegid.2012.02.002.
13. Sheen P, Couvin D, Grandjean L, Zimic M, Dominguez M, Luna G, et al. Genetic diversity of *Mycobacterium tuberculosis* in Peru and exploration of phylogenetic associations with drug resistance. *PloS One*. 2013;8(6):e65873. doi: 10.1371/journal.pone.0065873.
14. Grandjean L, Iwamoto T, Lithgow A, Gilman RH, Arikawa K, Nakanishi N, et al. The Association between *Mycobacterium Tuberculosis* Genotype and Drug Resistance in Peru. *PloS One*. 2015;10(5):e0126271. doi: 10.1371/journal.pone.0126271.



15. Tarazona D, Borda V, Galarza M, Agapito JC, Guio H. Functional Analysis Using Whole-Genome Sequencing of a Drug-Sensitive *Mycobacterium tuberculosis* Strain from Peru. *Genome Announc.* 2014;2(1). pii: e00087-14. doi: 10.1128/genomeA.00087-14.
16. Galarza M, Tarazona D, Borda V, Agapito JC, Guio H. Evidence of Clonal Expansion in the Genome of a Multidrug-Resistant *Mycobacterium tuberculosis* Clinical Isolate from Peru. *Genome Announc.* 2014;2(1). pii: e00089-14. doi: 10.1128/genomeA.00089-14.
17. Guio H, Tarazona D, Galarza M, Borda V, Curitomay R. Genome analysis of 17 extensively drug-resistant strains reveals new potential mutations for resistance. *Genome Announc.* 2014;2(4). pii: e00759-14. doi: 10.1128/genomeA.00759-14.
18. Stavrum R, Valvatne H, Stavrum AK, Riley LW, Ulvestad E, Jonassen I, et al. *Mycobacterium tuberculosis* Mce1 protein complex initiates rapid induction of transcription of genes involved in substrate trafficking. *Genes immun.* 2012;13(6):496-502. doi: 10.1038/gene.2012.24. Epub 2012 Jun 14.
19. Chandolia A, Rathor N, Sharma M, Saini NK, Sinha R, Malhotra P, et al. Functional analysis of mce4A gene of *Mycobacterium tuberculosis* H37Rv using antisense approach. *Microbiol Res.* 2014;169(9-10):780-7. doi: 10.1016/j.micres.2013.12.008.
20. Senaratne RH, Sidders B, Sequeira P, Saunders G, Dunphy K, Marjanovic O, et al. *Mycobacterium tuberculosis* strains disrupted in mce3 and mce4 operons are attenuated in mice. *J Med Microbiol.* 2008;57(Pt 2):164-70. doi: 10.1099/jmm.0.47454-0.
21. Bibb LA, Hatfull GF. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. *Mol Microbiol.* 2002;45(6):1515-26.
22. Fan X, Abd Alla AA, Xie J. Distribution and function of prophage phiRv1 and phiRv2 among *Mycobacterium tuberculosis* complex. *J Biomol Struct Dyn.* 2016;34(2):233-8. doi: 10.1080/07391102.2015.1022602.
23. Perez-Lago L, Navarro Y, Herranz M, Bouza E, Garcia-de-Viedma D. Genetic features shared by *Mycobacterium tuberculosis* strains involved in microevolution events. *Infect Genet Evol.* 2013;16:326-9. doi: 10.1016/j.meegid.2013.02.016.
24. Zheng H, Lu L, Wang B, Pu S, Zhang X, Zhu G, et al. Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PloS One.* 2008;3(6):e2375. doi: 10.1371/journal.pone.0002375.
25. Koh KW, Soh SE, Seah GT. Strong antibody responses to *Mycobacterium tuberculosis* PE-PGRS62 protein are associated with latent and active tuberculosis. *Infect Immun.* 2009;77(8):3337-43. doi: 10.1128/IAI.01175-08.
26. Ahmed A, Das A, Mukhopadhyay S. Immunoregulatory functions and expression patterns of PE/PPE family members: Roles in pathogenicity and impact on anti-tuberculosis vaccine and drug design. *IUBMB Life.* 2015;67(6):414-27. doi: 10.1002/iub.1387.
27. Smith KL, Saini D, Bardarov S, Larsen M, Frothingham R, Gandhi NR, et al. Reduced virulence of an extensively drug-resistant outbreak strain of *Mycobacterium tuberculosis* in a murine model. *PloS One.* 2014;9(4):e94953. doi: 10.1371/journal.pone.0094953.
28. Miggiano R, Casazza V, Garavaglia S, Ciaramella M, Perugino G, Rizzi M, et al. Biochemical and structural studies of the *Mycobacterium tuberculosis* O6-methylguanine

methyltransferase and mutated variants. J Bacteriol. 2013;195(12):2728-36. doi: 10.1128/JB.02298-12.

29. Rombouts Y, Brust B, Ojha AK, Maes E, Coddeville B, Ellass-Rochard E, et al. Exposure of mycobacteria to cell wall-inhibitory drugs decreases production of arabinoglycerolipid related to Mycolyl-arabinogalactan-peptidoglycan metabolism. J Biol Chem. 2012;287(14):11060-9. doi: 10.1074/jbc.M111.327387.

30. Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney JD, Mizrahi V, et al. A vitamin B12 transporter in Mycobacterium tuberculosis. Open Biol. 2013;3(2):120175. doi: 10.1098/rsob.120175.

**Correspondencia:**

David Tarazona

**Dirección:** Av. Defensores del Morro 2268, Chorrillos. Lima, Perú.

**Teléfono:** (+511) 7480000 anexo 1424

**Correo electrónico:** [ddtarazona@gmail.com](mailto:ddtarazona@gmail.com)

**Recibido:** 20/10/2015

**Aprobado:** 09/03/2016



Tabla Anexo 5. Variantes genéticas de *Mycobacterium tuberculosis* INS\_SEN

Posicion	GeneID	Producto	COG
1977	IGS	intergenic space or other non-protein-coding region	#N/A
3126	Rv0002	DNA polymerase III subunit beta	Replication, recombination and repair
4013	Rv0003	DNA replication and repair protein RecF	Replication, recombination and repair
4771	Rv0004	hypothetical protein	General function prediction only
7362	Rv0006	DNA gyrase subunit A	Replication, recombination and repair
7585	Rv0006	DNA gyrase subunit A	Replication, recombination and repair
9304	Rv0006	DNA gyrase subunit A	Replication, recombination and repair
11879	Rv0008c	membrane protein	#N/A
13035	IGS	intergenic space or other non-protein-coding region	#N/A
14251	Rv0012	hypothetical protein	Function unknown
20036	Rv0016c	penicillin-binding protein A	Cell wall/membrane/envelope biogenesis
21795	Rv0018c	PP2C-family Ser/Thr phosphatase	Signal transduction mechanisms
26959	IGS	intergenic space or other non-protein-coding region	#N/A
33457	Rv0030	hypothetical protein	#N/A
33551	Rv0030	hypothetical protein	#N/A
34044	IGS	intergenic space or other non-protein-coding region	#N/A
37031	Rv0034	hypothetical protein	General function prediction only
42967	Rv0040c	secreted proline rich protein mtc28	#N/A
47120	Rv0042c	MarR family transcriptional regulator	Transcription
47848	Rv0043c	GntR family transcriptional regulator	Transcription
62046	Rv0058	replicative DNA helicase	Replication, recombination and repair
63768	Rv0059	hypothetical protein	#N/A
64124	Rv0060	hypothetical protein	General function prediction only
66771	IGS	intergenic space or other non-protein-coding region	#N/A
70813	Rv0064	hypothetical protein	Function unknown
71333	Rv0064	hypothetical protein	Function unknown
71619	Rv0064A	antitoxin	#N/A
74056	Rv0066c	isocitrate dehydrogenase	Energy production and conversion
75937	Rv0068	oxidoreductase	Lipid transport and metabolism
76939	Rv0069c	L-serine dehydratase	Amino acid transport and metabolism
80613	IGS	intergenic space or other non-protein-coding region	#N/A
90474	Rv0083	oxidoreductase	Energy production and conversion
92196	Rv0083	oxidoreductase	Energy production and conversion
103753	Rv0094c	hypothetical protein	#N/A
105042	Rv0095c	hypothetical protein	#N/A

122106	Rv0103c	cation-transporting P-type ATPase B	Inorganic ion transport and metabolism
127173	Rv0107c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
133837	IGS	intergenic space or other non-protein-coding region	#N/A
133860	IGS	intergenic space or other non-protein-coding region	#N/A
146085	Rv0120c	elongation factor G-like protein	Translation, ribosomal structure and biogenesis
154136	IGS	intergenic space or other non-protein-coding region	#N/A
154281	Rv0127	hypothetical protein	Carbohydrate transport and metabolism
157290	Rv0129c	antigen 85-C	General function prediction only
162579	Rv0134	epoxide hydrolase EphF	General function prediction only
175419	Rv0148	oxidoreductase	Lipid transport and metabolism
177855	Rv0151c	PE family protein	#N/A
179961	Rv0152c	PE family protein	#N/A
180023	Rv0152c	PE family protein	#N/A
180371	Rv0152c	PE family protein	#N/A
188798	Rv0159c	PE family protein	#N/A
194679	Rv0165c	GntR family transcriptional regulator	#N/A
195211	Rv0166	fatty-acyl-CoA synthase	Lipid transport and metabolism
196640	Rv0166	fatty-acyl-CoA synthase	Lipid transport and metabolism
203267	Rv0172	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
206337	Rv0174	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
207224	Rv0175	MCE-associated membrane protein	#N/A
212351	Rv0181c	hypothetical protein	General function prediction only
217300	Rv0186	beta-glucosidase	Carbohydrate transport and metabolism
223940	Rv0192	hypothetical protein	Function unknown
225321	Rv0193c	hypothetical protein	#N/A
227096	Rv0194	ATP-binding cassette, subfamily C	Defense mechanisms
231112	Rv0195	two component system transcriptional regulator, luxR-family	Signal transduction mechanisms
246239	Rv0206c	transmembrane transporter mmpL3	General function prediction only
257374	Rv0214	long chain acyl-CoA synthetase	Lipid transport and metabolism
261869	Rv0218	hypothetical protein	General function prediction only
265554	Rv0222	enoyl-CoA hydratase	Lipid transport and metabolism
272513	Rv0227c	hypothetical protein	#N/A
278681	Rv0233	R2-like ligand binding oxidase	Nucleotide transport and metabolism
283614	Rv0236c	hypothetical protein	#N/A
285772	Rv0236c	hypothetical protein	#N/A
285871	Rv0236c	hypothetical protein	#N/A
302382	Rv0251c	heat shock protein hsp	Post-translational modification, protein turnover, and chaperones
304752	Rv0252	nitrite reductase [NAD(P)H], large subunit	Energy production and conversion

310973	Rv0259c	hypothetical protein	Function unknown
311613	Rv0260c	uroporphyrinogen-III synthase	Coenzyme transport and metabolism
316129	Rv0264c	hypothetical protein	Amino acid transport and metabolism
320900	Rv0266c	5-oxoprolinase	Amino acid transport and metabolism
325776	Rv0270	fatty-acyl-CoA synthase	Lipid transport and metabolism
333637	Rv0278c	PE-PGRS family protein	#N/A
333640	Rv0278c	PE-PGRS family protein	#N/A
333677	Rv0278c	PE-PGRS family protein	#N/A
333683	Rv0278c	PE-PGRS family protein	#N/A
333726	Rv0278c	PE-PGRS family protein	#N/A
333747	Rv0278c	PE-PGRS family protein	#N/A
333773	Rv0278c	PE-PGRS family protein	#N/A
333786	Rv0278c	PE-PGRS family protein	#N/A
333803	Rv0278c	PE-PGRS family protein	#N/A
333810	Rv0278c	PE-PGRS family protein	#N/A
335971	Rv0278c	PE-PGRS family protein	#N/A
336692	Rv0279c	WAG22 antigen	#N/A
337820	Rv0279c	WAG22 antigen	#N/A
340372	Rv0280	PPE family protein	Cell motility
344033	Rv0283	ESX-3 secretion system protein eccB3	#N/A
346275	Rv0284	ESX-3 secretion system protein eccC3	Cell cycle control, cell division, chromosome partitioning
356528	Rv0292	ESX-3 secretion system protein EccE3	#N/A
362013	Rv0297	PE-PGRS family protein	#N/A
362182	Rv0297	PE-PGRS family protein	#N/A
362188	Rv0297	PE-PGRS family protein	#N/A
366917	Rv0304c	PPE family protein	Cell motility
366922	Rv0304c	PPE family protein	Cell motility
366940	Rv0304c	PPE family protein	Cell motility
366946	Rv0304c	PPE family protein	Cell motility
373096	Rv0305c	PPE family protein	Cell motility
373934	Rv0305c	PPE family protein	Cell motility
376774	Rv0307c	hypothetical protein	#N/A
384380	Rv0315	hypothetical protein	Carbohydrate transport and metabolism
386432	Rv0318c	ZIP family zinc transporter	Inorganic ion transport and metabolism
403980	Rv0338c	oxidoreductase	Energy production and conversion
404326	Rv0338c	oxidoreductase	Energy production and conversion
408723	Rv0340	hypothetical protein	#N/A
414486	Rv0344c	lipoprotein LpqJ	Lipid transport and metabolism
420008	Rv0350	chaperone dnaK	Post-translational modification, protein turnover, and chaperones
425343	Rv0355c	PPE family protein	#N/A
427159	Rv0355c	PPE family protein	#N/A
427175	Rv0355c	PPE family protein	#N/A

444352	Rv0366c	hypothetical protein	Function unknown
454296	Rv0376c	xanthine dehydrogenase accessory factor	Post-translational modification, protein turnover, and chaperones
457453	Rv0381c	hypothetical protein	#N/A
458430	IGS	intergenic space or other non-protein-coding region	#N/A
472708	IGS	intergenic space or other non-protein-coding region	#N/A
475181	Rv0395	hypothetical protein	#N/A
488781	Rv0405	polyketide synthase 12	Secondary metabolites biosynthesis, transport, and catabolism
489938	Rv0405/Rv0406c	polyketide synthase 12/beta lactamase like protein	#N/A
498560	Rv0412c	hypothetical protein	#N/A
502592	Rv0417	thiazole synthase	Coenzyme transport and metabolism
503357	IGS	intergenic space or other non-protein-coding region	#N/A
513260	Rv0425c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
534694	Rv0446c	hypothetical protein	Function unknown
541204	Rv0450c	transmembrane transporter mmpL4	General function prediction only
551528	Rv0459	hypothetical protein	Function unknown
563423	Rv0472c	TetR family transcriptional regulator	Transcription
569956	IGS	intergenic space or other non-protein-coding region	#N/A
573265	Rv0484c	oxidoreductase	General function prediction only
584441	Rv0493c	hypothetical protein	#N/A
587721	Rv0497	hypothetical protein	#N/A
589539	Rv0499	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
590439	Rv0500	pyrroline-5-carboxylate reductase	Amino acid transport and metabolism
594648	Rv0503c	cyclopropane mycolic acid synthase 2	Cell wall/membrane/envelope biogenesis
596575	Rv0505c	HAD hydrolase, family IB	Amino acid transport and metabolism
597819	Rv0507	transmembrane transporter mmpL2	General function prediction only
598478	Rv0507	transmembrane transporter mmpL2	General function prediction only
610123	IGS	intergenic space or other non-protein-coding region	#N/A
615617	Rv0524	glutamate-1-semialdehyde 2,1-aminomutase	Coenzyme transport and metabolism
630725	Rv0538	hypothetical protein	General function prediction only
635636	Rv0543c	hypothetical protein	#N/A
637322	Rv0545c	PiT family inorganic phosphate transporter	Inorganic ion transport and metabolism
637925	Rv0546c	lactoylglutathione lyase	General function prediction only
648005	Rv0556	hypothetical protein	#N/A
662587	Rv0570	ribonucleoside-diphosphate reductase nrdZ	Nucleotide transport and metabolism
664252	Rv0571c	hypothetical protein	General function prediction only
665296	Rv0572c	hypothetical protein	#N/A
669401	Rv0575c	oxidoreductase	Coenzyme transport and metabolism

673241	Rv0578c	PE-PGRS family protein	#N/A
674955	Rv0578c	PE-PGRS family protein	#N/A
681619	Rv0584	hypothetical protein	Carbohydrate transport and metabolism
685464	Rv0587	ABC transporter permease	Secondary metabolites biosynthesis, transport, and catabolism
685611	Rv0587	ABC transporter permease	Secondary metabolites biosynthesis, transport, and catabolism
686975	Rv0589	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
690468	Rv0591	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
691312	Rv0592	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
696421	Rv0597c	ATPase AAA+ superfamily protein	General function prediction only
697554	Rv0598c	toxin	#N/A
698971	Rv0601c	sensor histidine kinase component HK2	Signal transduction mechanisms
709229	Rv0613c	hypothetical protein	#N/A
718439	Rv0627	toxin	General function prediction only
726706	Rv0631c	exodeoxyribonuclease V subunit gamma	Replication, recombination and repair
752346	Rv0655	ABC transporter ATP-binding protein	Secondary metabolites biosynthesis, transport, and catabolism
754189	Rv0658c	hypothetical protein	General function prediction only
758258	Rv0663	arylsulfatase	Inorganic ion transport and metabolism
764998	Rv0668	DNA-directed RNA polymerase subunit beta'	Transcription
769257	Rv0669c	hydrolase	#N/A
773812	Rv0673	enoyl-CoA hydratase	Lipid transport and metabolism
775642	Rv0676c	transmembrane transporter mmpL5	General function prediction only
781398	IGS	intergenic space or other non-protein-coding region	#N/A
786510	Rv0686	membrane protein	#N/A
796512	Rv0696	mycofactocin system glycosyltransferase	Cell wall/membrane/envelope biogenesis
820486	Rv0727c	L-ribulose-5-phosphate 4-epimerase	Carbohydrate transport and metabolism
826178	Rv0733	adenylate kinase	Nucleotide transport and metabolism
839952	Rv0747	PE-PGRS family protein	#N/A
841767	Rv0749A	hypothetical protein	#N/A
845673	Rv0753c	methyilmalonate-semialdehyde dehydrogenase	Energy production and conversion
849894	Rv0755c	PPE family protein	Cell motility
851195	Rv0756c	hypothetical protein	#N/A
852913	Rv0758	two-component system, OmpR family, sensor kinase	Signal transduction mechanisms
857699	Rv0764c	lanosterol 14-alpha demethylase	Secondary metabolites biosynthesis, transport, and catabolism
880565	Rv0785	KsdD-like steroid dehydrogenase	General function prediction only

882260	Rv0787	hypothetical protein	#N/A
885545	Rv0791c	hypothetical protein	Energy production and conversion
893736	Rv0800	aminopeptidase	Amino acid transport and metabolism
894871	Rv0801	hypothetical protein	General function prediction only
900224	Rv0806c	exopolysaccharide phosphotransferase CpsY	#N/A
903553	Rv0808	amidophosphoribosyltransferase	Nucleotide transport and metabolism
903916	Rv0809	phosphoribosylformylglycinamide cyclo-ligase	Nucleotide transport and metabolism
906860	Rv0812	4-amino-4-deoxychorismate lyase	Amino acid transport and metabolism
927388	Rv0833	PE-PGRS family protein	Function unknown
942482	IGS	intergenic space or other non-protein-coding region	#N/A
942597	IGS	intergenic space or other non-protein-coding region	#N/A
945217	Rv0848	cysteine synthase A	Amino acid transport and metabolism
947432	Rv0850	transposase	#N/A
949538	Rv0853c	alpha-keto-acid decarboxylase	Carbohydrate transport and metabolism
955527	Rv0859	acetyl-CoA C-acetyltransferase	Lipid transport and metabolism
958925	Rv0861c	DNA excision repair protein ERCC-3	#N/A
966945	Rv0869c	molybdenum cofactor biosynthesis protein A 2	Coenzyme transport and metabolism
979707	Rv0881	methytransferase	Translation, ribosomal structure and biogenesis
984264	Rv0886	ferredoxin-NADP+ reductase	Amino acid transport and metabolism
986466	IGS	intergenic space or other non-protein-coding region	#N/A
993349	Rv0891c	transcriptional regulator	Signal transduction mechanisms
1037016	Rv0930	phosphate transporter permease pstA 1	Inorganic ion transport and metabolism
1041589	Rv0933	phosphate import ATP-binding protein PstB 2	Inorganic ion transport and metabolism
1047169	Rv0938	DNA ligase D	Replication, recombination and repair
1068155	Rv0956	phosphoribosylglycinamide formyltransferase	Nucleotide transport and metabolism
1068436	Rv0957	bifunctional purine biosynthesis protein purH	Nucleotide transport and metabolism
1070706	Rv0958	magnesium chelatase subunit I	Coenzyme transport and metabolism
1071179	Rv0958	magnesium chelatase subunit I	Coenzyme transport and metabolism
1074562	Rv0962c	lipoprotein LprP	#N/A
1075283	IGS	intergenic space or other non-protein-coding region	#N/A
1076313	Rv0964c	hypothetical protein	#N/A
1077316	Rv0966c	hypothetical protein	#N/A
1079931	Rv0969	heavy metal translocating P-type ATPase	#N/A
1081685	Rv0970	hypothetical protein	#N/A
1087197	Rv0974c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
1089439	Rv0976c	hypothetical protein	#N/A
1093326	IGS	intergenic space or other non-protein-coding region	#N/A

1096636	IGS	intergenic space or other non-protein-coding region	#N/A
1096952	Rv0981	response regulator mprA	#N/A
1096953	Rv0981	response regulator mprA	#N/A
1100238	Rv0983	serine protease PepD	Post-translational modification, protein turnover, and chaperones
1103253	Rv0987	ABC transporter permease	Cell wall/membrane/envelope biogenesis
1106426	Rv0989c	heptaprenyl diphosphate synthase	Coenzyme transport and metabolism
1107438	IGS	intergenic space or other non-protein-coding region	#N/A
1107921	Rv0990c	hypothetical protein	Cell motility
1109979	Rv0993	UTP-glucose-1-phosphate uridylyltransferase	Cell wall/membrane/envelope biogenesis
1117581	Rv1001	arginine deiminase	Amino acid transport and metabolism
1120744	Rv1003	ribosomal RNA small subunit methyltransferase I	General function prediction only
1126893	Rv1007c	methionyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1127652	Rv1008	TatD family hydrolase	Replication, recombination and repair
1142270	Rv1020	transcription-repair-coupling factor	Replication, recombination and repair
1149555	Rv1028c	sensor protein kdpD	Signal transduction mechanisms
1150589	Rv1028c	sensor protein kdpD	Signal transduction mechanisms
1155655	Rv1030	potassium-transporting ATPase B chain	Inorganic ion transport and metabolism
1163138	Rv1040c	PE family protein	#N/A
1172869	IGS	intergenic space or other non-protein-coding region	#N/A
1174723	IGS	intergenic space or other non-protein-coding region	#N/A
1178121	Rv1056	hypothetical protein	Function unknown
1190177	Rv1067c	PE-PGRS family protein	Function unknown
1199552	Rv1075c	hypothetical protein	Amino acid transport and metabolism
1200423	IGS	intergenic space or other non-protein-coding region	#N/A
1201823	Rv1077	cystathionine beta-synthase	Amino acid transport and metabolism
1202111	Rv1077	cystathionine beta-synthase	Amino acid transport and metabolism
1203602	Rv1078	proline-rich antigen pra	RNA processing and modification
1209536	Rv1084	hypothetical protein	Post-translational modification, protein turnover, and chaperones
1210860	Rv1086	short-chain Z-isoprenyl diphosphate synthase	Lipid transport and metabolism
1212081	Rv1087	PE-PGRS family protein	Function unknown
1218742	Rv1091	PE-PGRS family protein	#N/A
1220685	Rv1093	serine hydroxymethyltransferase 1	#N/A
1224372	IGS	intergenic space or other non-protein-coding region	#N/A
1248983	Rv1125	hypothetical protein	#N/A
1281123	Rv1154c	hypothetical protein	Function unknown
1292107	Rv1162	nitrate reductase subunit beta	Energy production and conversion
1302448	Rv1172c	PE family protein	#N/A

1307603	Rv1175c	2,4-dienoyl-CoA reductase	Energy production and conversion
1327896	Rv1186c	hypothetical protein	Signal transduction mechanisms
1328693	Rv1186c	hypothetical protein	Signal transduction mechanisms
1346528	Rv1203c	hypothetical protein	#N/A
1354443	Rv1211	hypothetical protein	#N/A
1360215	Rv1217c	antibiotic transporter permease	Cell wall/membrane/envelope biogenesis
1370334	Rv1227c	transmembrane protein	Function unknown
1373672	Rv1230c	membrane protein	Cell wall/membrane/envelope biogenesis
1374071	Rv1230c	membrane protein	Cell wall/membrane/envelope biogenesis
1375730	Rv1232c	hypothetical protein	Inorganic ion transport and metabolism
1382634	Rv1239c	magnesium and cobalt transporter CorA	Inorganic ion transport and metabolism
1389744	Rv1248c	2-oxoglutarate decarboxylase	Energy production and conversion
1393632	Rv1249c	membrane protein	#N/A
1396928	Rv1251c	hypothetical protein	General function prediction only
1404175	Rv1256c	cytochrome P450 130 cyp130	Secondary metabolites biosynthesis, transport, and catabolism
1411216	Rv1263	amidase	Translation, ribosomal structure and biogenesis
1413154	IGS	intergenic space or other non-protein-coding region	#N/A
1427902	Rv1278	hypothetical protein	Function unknown
1428512	Rv1278	hypothetical protein	Function unknown
1440475	Rv1286	bifunctional enzyme CysN/CysC	Inorganic ion transport and metabolism
1445480	IGS	intergenic space or other non-protein-coding region	#N/A
1445787	Rv1291c	hypothetical protein	#N/A
1446929	Rv1292	arginyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1452077	Rv1296	homoserine kinase	Amino acid transport and metabolism
1457150	Rv1300	HemK family methyltransferase	Translation, ribosomal structure and biogenesis
1458198	Rv1301	Sua5/YciO/YrdC/YwIc family protein	Translation, ribosomal structure and biogenesis
1480030	Rv1318c	adenylate cyclase	Signal transduction mechanisms
1482633	Rv1320c	adenylate cyclase	Signal transduction mechanisms
1484714	Rv1321	hypothetical protein	Replication, recombination and repair
1487760	Rv1324	thioredoxin	Post-translational modification, protein turnover, and chaperones
1496301	Rv1328	glycogen phosphorylase	Carbohydrate transport and metabolism
1499280	Rv1330c	nicotinate phosphoribosyltransferase pncB1	Coenzyme transport and metabolism
1503754	Rv1336	O-phosphoserine sulfhydrylase	Amino acid transport and metabolism
1511805	Rv1346	acyl-[acyl-carrier-protein] dehydrogenase MbtN	Lipid transport and metabolism
1512037	Rv1347c	lysine N-acyltransferase mbtK	Translation, ribosomal structure and biogenesis
1526825	Rv1358	transcriptional regulator	General function prediction only
1532662	Rv1361c	PPE family protein	Cell motility



1536257	Rv1364c	anti-anti-sigma factor	Signal transduction mechanisms
1543481	Rv1371	hypothetical protein	Lipid transport and metabolism
1547131	Rv1374c	hypothetical protein	#N/A
1552553	Rv1378c	hypothetical protein	#N/A
1563723	Rv1388	integration host factor mihF	#N/A
1570199	Rv1394c	cytochrome P450 132 cyp132	Secondary metabolites biosynthesis, transport, and catabolism
1570572	Rv1394c	cytochrome P450 132 cyp132	Secondary metabolites biosynthesis, transport, and catabolism
1572206	Rv1396c	PE-PGRS family protein	Cell wall/membrane/envelope biogenesis
1573666	Rv1396c	PE-PGRS family protein	Cell wall/membrane/envelope biogenesis
1586255	Rv1410c	MFS-type drug efflux transporter P55	Carbohydrate transport and metabolism
1588905	Rv1412	riboflavin synthase subunit alpha	Coenzyme transport and metabolism
1609846	Rv1431	hypothetical protein	#N/A
1613041	IGS	intergenic space or other non-protein-coding region	#N/A
1613966	Rv1436	glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate transport and metabolism
1624797	Rv1446c	opcA protein	Carbohydrate transport and metabolism
1625065	Rv1446c	opcA protein	Carbohydrate transport and metabolism
1630154	Rv1449c	transketolase	Carbohydrate transport and metabolism
1632913	Rv1450c	PE-PGRS family protein	#N/A
1636832	Rv1452c	PE-PGRS family protein	#N/A
1636924	Rv1452c	PE-PGRS family protein	#N/A
1636933	Rv1452c	PE-PGRS family protein	#N/A
1636940	Rv1452c	PE-PGRS family protein	#N/A
1636942	Rv1452c	PE-PGRS family protein	#N/A
1639600	Rv1453	hypothetical protein	Signal transduction mechanisms
1645808	Rv1459c	hypothetical protein	#N/A
1650078	Rv1462	Fe-S cluster assembly protein SufD	Post-translational modification, protein turnover, and chaperones
1669260	IGS	intergenic space or other non-protein-coding region	#N/A
1676296	Rv1486c	hypothetical protein	#N/A
1688013	Rv1497	esterase lipL	Defense mechanisms
1689355	Rv1498c	methyltransferase	Coenzyme transport and metabolism
1690832	IGS	intergenic space or other non-protein-coding region	#N/A
1692147	Rv1501	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
1692801	IGS	intergenic space or other non-protein-coding region	#N/A
1693567	Rv1502	hypothetical protein	#N/A
1698098	IGS	intergenic space or other non-protein-coding region	#N/A
1708563	IGS	intergenic space or other non-protein-coding region	#N/A

1711633	Rv1520	sugar transferase	#N/A
1718767	Rv1524	glycosyltransferase	Carbohydrate transport and metabolism
1722842	Rv1527c	phthioceranic/hydroxyphthioceranic acid synthase	Secondary metabolites biosynthesis, transport, and catabolism
1728843	IGS	intergenic space or other non-protein-coding region	#N/A
1735000	Rv1534	transcriptional regulator	Transcription
1740777	Rv1537	DNA polymerase IV 1	#N/A
1752567	Rv1548c	PPE family protein	Cell motility
1759258	Rv1552	fumarate reductase flavoprotein subunit	Energy production and conversion
1760298	Rv1554	fumarate reductase subunit C	Energy production and conversion
1778436	Rv1570	dethiobiotin synthetase	Coenzyme transport and metabolism
1779285	Rv1572c	hypothetical protein	#N/A
1782625	IGS	intergenic space or other non-protein-coding region	#N/A
1782717	IGS	intergenic space or other non-protein-coding region	#N/A
1782774	Rv1578c	phiRv1 phage protein	Replication, recombination and repair
1782780	Rv1578c	phiRv1 phage protein	Replication, recombination and repair
1782791	Rv1578c	phiRv1 phage protein	Replication, recombination and repair
1782793	Rv1578c	phiRv1 phage protein	Replication, recombination and repair
1782799	Rv1578c	phiRv1 phage protein	Replication, recombination and repair
1793776	IGS	intergenic space or other non-protein-coding region	#N/A
1798331	Rv1597	hypothetical protein	Coenzyme transport and metabolism
1798362	Rv1597	hypothetical protein	Coenzyme transport and metabolism
1803272	Rv1602	imidazole glycerol phosphate synthase subunit HisH	Amino acid transport and metabolism
1804416	Rv1604	myo-inositol-1(or 4)-monophosphatase	Carbohydrate transport and metabolism
1809969	Rv1610	hypothetical protein	#N/A
1817983	Rv1618	acyl-CoA thioesterase II	Lipid transport and metabolism
1836293	IGS	intergenic space or other non-protein-coding region	#N/A
1847926	Rv1639c	hypothetical protein	General function prediction only
1854307	Rv1644	TrmH family RNA methyltransferase	Translation, ribosomal structure and biogenesis
1856784	Rv1647	adenylate cyclase	Signal transduction mechanisms
1863728	Rv1651c	PE-PGRS family protein	#N/A
1863844	Rv1651c	PE-PGRS family protein	#N/A
1878790	Rv1661	polyketide synthase 7	Secondary metabolites biosynthesis, transport, and catabolism
1885779	Rv1662	polyketide synthase 8	Secondary metabolites biosynthesis, transport, and catabolism
1899358	Rv1674c	transcriptional regulator	Inorganic ion transport and metabolism

1899462	Rv1674c	transcriptional regulator	Inorganic ion transport and metabolism
1900168	Rv1674A	hypothetical protein	#N/A
1901500	Rv1676	hypothetical protein	Post-translational modification, protein turnover, and chaperones
1907303	Rv1682	hypothetical protein	#N/A
1914638	Rv1689	tyrosyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1917979	Rv1694	hemolysin TlyA family protein	Translation, ribosomal structure and biogenesis
1920687	Rv1696	DNA repair protein recN	Replication, recombination and repair
1931186	Rv1704c	AAT family amino acid transporter	Amino acid transport and metabolism
1933995	Rv1706Xc	hypothetical protein	#N/A
1942891	Rv1715	3-hydroxybutyryl-CoA dehydrogenase	Lipid transport and metabolism
1943046	Rv1715	3-hydroxybutyryl-CoA dehydrogenase	Lipid transport and metabolism
1943599	Rv1716	arylformamidase	General function prediction only
1944114	Rv1716	arylformamidase	General function prediction only
1944409	Rv1716	arylformamidase	General function prediction only
1947910	Rv1722	carboxylase	Lipid transport and metabolism
1950774	Rv1724c	hypothetical protein	#N/A
1955643	IGS	intergenic space or other non-protein-coding region	#N/A
1960291	Rv1733c	hypothetical protein	#N/A
1965631	IGS	intergenic space or other non-protein-coding region	#N/A
1967244	Rv1739c	sulfate permease	Inorganic ion transport and metabolism
1981063	IGS	intergenic space or other non-protein-coding region	#N/A
1983320	Rv1753c	PPE family protein	Cell motility
1989066	Rv1758	esterase/lipase	#N/A
1989086	Rv1758	esterase/lipase	#N/A
1990926	Rv1759c	WAG22 antigen	#N/A
1990963	Rv1759c	WAG22 antigen	#N/A
1993815	Rv1760	acyltransferase, WS/DGAT/MGAT	#N/A
2006039	Rv1771	FAD-linked oxidoreductase	Energy production and conversion
2009080	Rv1774	oxidoreductase	Energy production and conversion
2014086	Rv1779c	hypothetical protein	#N/A
2022875	Rv1784	ESX-5 secretion system protein eccCa5	#N/A
2045317	Rv1803c	hypothetical protein	#N/A
2049072	Rv1807	hypothetical protein	Cell motility
2049104	Rv1807	hypothetical protein	Cell motility
2052042	Rv1809	hypothetical protein	Cell motility
2053418	IGS	intergenic space or other non-protein-coding region	#N/A
2053689	Rv1811	Mg2+ transporter-C (MgtC) family protein	Function unknown
2055278	Rv1812c	NADH dehydrogenase-like protein	Energy production and conversion
2074516	IGS	intergenic space or other non-protein-coding region	#N/A
2074572	IGS	intergenic space or other non-protein-coding region	#N/A

2079106	Rv1833c	haloalkane dehalogenase 2	General function prediction only
2096193	Rv1846c	transcriptional regulator Blal	Transcription
2097997	Rv1850	urease subunit alpha	Amino acid transport and metabolism
2108148	Rv1860	alanine and proline-rich secreted protein apa	Function unknown
2110357	Rv1862	zinc-binding alcohol dehydrogenase	General function prediction only
2116910	Rv1867	acetyl-CoA C-acetyltransferase	Lipid transport and metabolism
2123176	IGS	intergenic space or other non-protein-coding region	#N/A
2128877	Rv1878	glutamine synthetase	Amino acid transport and metabolism
2135877	IGS	intergenic space or other non-protein-coding region	#N/A
2143335	Rv1895	chlorophyll synthesis pathway, bchC	Amino acid transport and metabolism
2147029	Rv1900c	hypothetical protein	Signal transduction mechanisms
2153417	Rv1907c	hypothetical protein	#N/A
2163797	Rv1917c	hypothetical protein	Cell motility
2165293	Rv1917c	hypothetical protein	Cell motility
2165561	Rv1917c	hypothetical protein	Cell motility
2174223	Rv1922	hypothetical protein	Defense mechanisms
2199848	Rv1949c	hypothetical protein	#N/A
2202401	Rv1956	antitoxin, transcriptional regulator	Transcription
2203031	Rv1957/Rv1958c	hypothetical protein//hypothetical protein	#N/A
2207532	IGS	intergenic space or other non-protein-coding region	#N/A
2211834	Rv1968	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2213551	Rv1969	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2216256	Rv1971	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2216451	Rv1971	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2220520	Rv1977	hypothetical protein	Post-translational modification, protein turnover, and chaperones
2223301	IGS	intergenic space or other non-protein-coding region	#N/A
2228975	IGS	intergenic space or other non-protein-coding region	#N/A
2231140	Rv1987	chitinase	Carbohydrate transport and metabolism
2231527	IGS	intergenic space or other non-protein-coding region	#N/A
2245540	Rv2000	hypothetical protein	Energy production and conversion
2254065	Rv2006	trehalose-phosphatase	Carbohydrate transport and metabolism
2260533	IGS	intergenic space or other non-protein-coding region	#N/A
2264790	Rv2017	transcriptional regulator	Amino acid transport and metabolism
2265067	IGS	intergenic space or other non-protein-coding region	#N/A
2269788	Rv2024c	hypothetical protein	General function prediction only

2270110	Rv2024c	hypothetical protein	General function prediction only
2281308	Rv2034	ArsR family transcriptional regulator	Transcription
2283038	Rv2037c	hypothetical protein	General function prediction only
2285259	Rv2039c	multiple sugar transporter permease	Carbohydrate transport and metabolism
2287129	Rv2041c	multiple sugar transporter substrate-binding protein	Carbohydrate transport and metabolism
2292146	Rv2047c	hypothetical protein	Cell wall/membrane/envelope biogenesis
2300245	Rv2048c	polyketide synthase 12	#N/A
2300554	Rv2048c	polyketide synthase 12	#N/A
2300560	Rv2048c	polyketide synthase 12	#N/A
2300563	Rv2048c	polyketide synthase 12	#N/A
2306314	Rv2048c	polyketide synthase 12	#N/A
2326897	IGS	intergenic space or other non-protein-coding region	#N/A
2329541	Rv2072c	precorrin-6Y C(5,15)-methyltransferase	Coenzyme transport and metabolism
2334015	Rv2077c	hypothetical protein	#N/A
2335083	Rv2078	hypothetical protein	#N/A
2335502	Rv2079	hypothetical protein	Cell wall/membrane/envelope biogenesis
2340629	Rv2082	hypothetical protein	#N/A
2341644	Rv2083	hypothetical protein	#N/A
2345045	Rv2088	serine/threonine protein kinase	General function prediction only
2348454	Rv2090	5'-3' exonuclease	Replication, recombination and repair
2361612	Rv2101	helicase helZ	#N/A
2362049	Rv2101	helicase helZ	#N/A
2362296	Rv2101	helicase helZ	#N/A
2386397	Rv2125	hypothetical protein	General function prediction only
2415664	Rv2155c	UDP-N-acetylmuramoylalanine--D-glutamate ligase	#N/A
2424933	IGS	intergenic space or other non-protein-coding region	#N/A
2425670	Rv2163c	cell division protein FtsI (penicillin-binding protein 3)	Cell wall/membrane/envelope biogenesis
2440934	Rv2178c	3-deoxy-7-phosphoheptulonate synthase	Amino acid transport and metabolism
2453610	Rv2190Ac	hypothetical protein	#N/A
2456535	Rv2192c	anthranilate phosphoribosyltransferase	Amino acid transport and metabolism
2462879	Rv2198c	membrane protein mmpS3	General function prediction only
2463102	IGS	intergenic space or other non-protein-coding region	#N/A
2469583	Rv2205c	glycerate kinase	Carbohydrate transport and metabolism
2484263	Rv2216	epimerase	General function prediction only
2499734	Rv2226	hypothetical protein	Function unknown
2502081	Rv2228c	hypothetical protein	Carbohydrate transport and metabolism
2503633	Rv2230c	hypothetical protein	Function unknown
2509148	Rv2236c	cobalamin biosynthesis protein CobD	Coenzyme transport and metabolism

2509325	Rv2236c	cobalamin biosynthesis protein CobD	Coenzyme transport and metabolism
2509730	Rv2237	hypothetical protein	Function unknown
2513473	Rv2241	pyruvate dehydrogenase E1 component	#N/A
2514818	Rv2241	pyruvate dehydrogenase E1 component	#N/A
2521350	Rv2247	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
2529690	Rv2256c	hypothetical protein	#N/A
2531752	Rv2258c	transcriptional regulator	Coenzyme transport and metabolism
2535308	Rv2262c	apolipoprotein N-acyltransferase	Cell wall/membrane/envelope biogenesis
2549703	Rv2277c	glycerolphosphodiesterase	Energy production and conversion
2573766	IGS	intergenic space or other non-protein-coding region	#N/A
2581065	Rv2308	hypothetical protein	Function unknown
2586137	Rv2314c	hypothetical protein	General function prediction only
2598410	Rv2326c	cobalt/nickel transporter ATP-binding protein	General function prediction only
2611598	Rv2336	hypothetical protein	#N/A
2612633	Rv2337c	hypothetical protein	#N/A
2612642	Rv2337c	hypothetical protein	#N/A
2613362	Rv2338c	molybdopterin biosynthesis protein MoeW	Coenzyme transport and metabolism
2621068	Rv2343c	DNA primase	Replication, recombination and repair
2654381	Rv2374c	heat-inducible transcription repressor hrcA	Transcription
2656235	Rv2377c	protein mbtH	Function unknown
2660329	Rv2379c	amino acid adenylation domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
2680668	Rv2386A	hypothetical protein	#N/A
2695388	Rv2398c	sulfate ABC transporter, permease CysW	Inorganic ion transport and metabolism
2713805	IGS	intergenic space or other non-protein-coding region	#N/A
2721753	Rv2424c	transposase	Replication, recombination and repair
2734084	Rv2436	ribokinase	Carbohydrate transport and metabolism
2736444	Rv2438c	glutamine-dependent NAD(+) synthetase	Coenzyme transport and metabolism
2736751	Rv2438c/Rv2438A	glutamine-dependent NAD(+) synthetase//hypothetical protein	#N/A
2745152	IGS	intergenic space or other non-protein-coding region	#N/A
2745899	Rv2446c	hypothetical protein	#N/A
2746986	Rv2447c	FolC protein	Coenzyme transport and metabolism
2752708	IGS	intergenic space or other non-protein-coding region	#N/A
2760162	Rv2458	homocysteine S-methyltransferase	Amino acid transport and metabolism
2765892	Rv2463	esterase/lipase lipP	Defense mechanisms
2779146	Rv2476c	glutamate dehydrogenase	Amino acid transport and metabolism
2785981	IGS	intergenic space or other non-protein-coding region	#N/A
2786962	Rv2482c	glycerol-3-phosphate acyltransferase	Lipid transport and metabolism

2792520	IGS	intergenic space or other non-protein-coding region	#N/A
2796312	Rv2487c	PE-PGRS family protein	#N/A
2818847	Rv2503c	3-oxoacid CoA-transferase subunit B	Lipid transport and metabolism
2821352	Rv2505c	fatty-acyl-CoA synthase	Lipid transport and metabolism
2827994	IGS	intergenic space or other non-protein-coding region	#N/A
2828029	IGS	intergenic space or other non-protein-coding region	#N/A
2830535	Rv2513	hypothetical protein	#N/A
2855269	Rv2531c	arginine decarboxylase	Amino acid transport and metabolism
2865770	Rv2542	hypothetical protein	Replication, recombination and repair
2865892	Rv2542	hypothetical protein	Replication, recombination and repair
2871058	Rv2551c	leader peptidase (prepilin peptidase)/N-methyltransferase	Cell motility
2880712	Rv2560	hypothetical protein	Function unknown
2881465	Rv2561	hypothetical protein	#N/A
2882337	Rv2563	ABC transporter permease	Cell wall/membrane/envelope biogenesis
2888211	Rv2566	hypothetical protein	Function unknown
2889643	Rv2566	hypothetical protein	Function unknown
2891277	Rv2567	hypothetical protein	Function unknown
2891738	Rv2567	hypothetical protein	Function unknown
2893248	Rv2568c	hypothetical protein	Function unknown
2894218	Rv2569c	hypothetical protein	Amino acid transport and metabolism
2910471	Rv2584c	adenine phosphoribosyltransferase	Nucleotide transport and metabolism
2911303	Rv2585c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
2912304	Rv2585c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
2923401	Rv2592c	Holliday junction DNA helicase subunit RuvB	Replication, recombination and repair
2927949	IGS	intergenic space or other non-protein-coding region	#N/A
2939383	Rv2611c	phosphatidylinositol mannoside acyltransferase	Cell wall/membrane/envelope biogenesis
2939667	Rv2611c	phosphatidylinositol mannoside acyltransferase	Cell wall/membrane/envelope biogenesis
2945177	IGS	intergenic space or other non-protein-coding region	#N/A
2968923	IGS	intergenic space or other non-protein-coding region	#N/A
2970945	Rv2646	hypothetical protein	Replication, recombination and repair
2973473	IGS	intergenic space or other non-protein-coding region	#N/A
2974943	Rv2650c	hypothetical protein	General function prediction only
2976574	IGS	intergenic space or other non-protein-coding region	#N/A
2984750	Rv2668	hypothetical protein	#N/A
3005195	Rv2688c	fluoroquinolones export ATP-binding protein	Defense mechanisms
3009702	Rv2691	trk system potassium uptake protein TrkA	Inorganic ion transport and metabolism

3017475	Rv2702	polyphosphate glucokinase	Transcription
3027698	Rv2714	hypothetical protein	#N/A
3041881	Rv2729c	hypothetical protein	#N/A
3042856	Rv2730	hypothetical protein	#N/A
3054091	Rv2741	PE-PGRS family protein	#N/A
3054331	Rv2741	PE-PGRS family protein	#N/A
3054430	Rv2741	PE-PGRS family protein	#N/A
3054439	Rv2741	PE-PGRS family protein	#N/A
3063375	Rv2750	SDR family oxidoreductase	Lipid transport and metabolism
3073878	Rv2764c	thymidylate synthase	Nucleotide transport and metabolism
3077049	Rv2768c	PPE family protein	Cell motility
3080805	Rv2771c	hypothetical protein	General function prediction only
3086798	IGS	intergenic space or other non-protein-coding region	#N/A
3103692	Rv2794c	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
3108065	Rv2799	hypothetical protein	#N/A
3108684	Rv2800	hydrolase	General function prediction only
3118010	Rv2812	transposase	Replication, recombination and repair
3121890	IGS	intergenic space or other non-protein-coding region	#N/A
3122033	Rv2815Ac	hypothetical protein	#N/A
3122059	Rv2815Ac	hypothetical protein	#N/A
3122070	Rv2815Ac	hypothetical protein	#N/A
3122086	Rv2815Ac	hypothetical protein	#N/A
3122093	Rv2815Ac	hypothetical protein	#N/A
3133546	Rv2825c	hypothetical protein	Function unknown
3136561	Rv2828Ac	hypothetical protein	#N/A
3137068	Rv2830c	antitoxin	Cell cycle control, cell division, chromosome partitioning
3148398	Rv2841c	transcription elongation protein nusA	Transcription
3162815	Rv2853	PE-PGRS family protein	Function unknown
3165084	Rv2854	hypothetical protein	Lipid transport and metabolism
3174361	Rv2862c	hypothetical protein	#N/A
3175712	Rv2864c	penicillin-binding lipoprotein	Cell wall/membrane/envelope biogenesis
3177894	Rv2866	toxin	Translation, ribosomal structure and biogenesis
3178455	IGS	intergenic space or other non-protein-coding region	#N/A
3186870	Rv2874	protein dipZ	Post-translational modification, protein turnover, and chaperones
3187545	Rv2875	immunogenic protein MPT70	Cell wall/membrane/envelope biogenesis
3191037	Rv2881c	phosphatidate cytidyltransferase	Lipid transport and metabolism
3226191	Rv2916c	signal recognition particle protein	Intracellular trafficking, secretion, and vesicular transport
3228153	Rv2917	hypothetical protein	Transcription
3232825	IGS	intergenic space or other non-protein-coding region	#N/A



3236240	Rv2922c	chromosome partition protein smc	Cell cycle control, cell division, chromosome partitioning
3243640	IGS	intergenic space or other non-protein-coding region	#N/A
3248709	Rv2931	phenolphthiocerol synthesis polyketide synthase PpsA	Secondary metabolites biosynthesis, transport, and catabolism
3256504	Rv2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3260311	Rv2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3260970	Rv2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3264364	Rv2934	phenolphthiocerol synthesis type-I polyketide synthase D	Secondary metabolites biosynthesis, transport, and catabolism
3269591	Rv2935	non-ribosomal peptide synthetase domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
3270794	Rv2935	non-ribosomal peptide synthetase domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
3292088	Rv2946c	polyketide synthase 1/15	Secondary metabolites biosynthesis, transport, and catabolism
3296853	Rv2947c	polyketide synthase 1/15	Secondary metabolites biosynthesis, transport, and catabolism
3300758	Rv2950c	long chain fatty acid CoA FadD26	#N/A
3304143	Rv2951c	phthiodylone/phenolphthiodylone dimycocerosates ketoreductase	Energy production and conversion
3308616	IGS	intergenic space or other non-protein-coding region	#N/A
3336597	IGS	intergenic space or other non-protein-coding region	#N/A
3336835	Rv2981c	D-alanine--D-alanine ligase	Cell wall/membrane/envelope biogenesis
3338613	Rv2982c	glycerol-3-phosphate dehydrogenase [NAD(P)+]	Energy production and conversion
3343941	IGS	intergenic space or other non-protein-coding region	#N/A
3354906	Rv2996c	D-3-phosphoglycerate dehydrogenase	Coenzyme transport and metabolism
3358245	Rv2999	lipoprotein LppY	#N/A
3363348	IGS	intergenic space or other non-protein-coding region	#N/A
3367775	Rv3009c	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	Translation, ribosomal structure and biogenesis
3374110	Rv3014c	DNA ligase	Replication, recombination and repair
3392949	Rv3032A	hypothetical protein	#N/A
3395520	Rv3035	hypothetical protein	Function unknown
3402827	Rv3042c	phosphoserine phosphatase	Amino acid transport and metabolism
3418325	Rv3057c	short-chain type alcohol dehydrogenase/reductase	General function prediction only
3418327	Rv3057c	short-chain type alcohol dehydrogenase/reductase	General function prediction only
3425851	Rv3062	DNA ligase I	Replication, recombination and repair

3426792	Rv3062	DNA ligase I	Replication, recombination and repair
3428914	Rv3063	carbon starvation protein	Signal transduction mechanisms
3429199	Rv3063	carbon starvation protein	Signal transduction mechanisms
3440461	Rv3077	hydrolase	Inorganic ion transport and metabolism
3440465	Rv3077	hydrolase	Inorganic ion transport and metabolism
3456663	Rv3089	o-succinylbenzoate--CoA ligase	Lipid transport and metabolism
3460983	Rv3092c	hypothetical protein	Function unknown
3462132	Rv3093c	oxidoreductase	Energy production and conversion
3465836	Rv3097c	triacylglycerol lipase	Lipid transport and metabolism
3486974	Rv3121	cytochrome P450 141 cyp141	Secondary metabolites biosynthesis, transport, and catabolism
3503892	Rv3137	histidinol-phosphatase	Carbohydrate transport and metabolism
3515570	Rv3150	NADH-quinone oxidoreductase subunit F	Energy production and conversion
3518164	Rv3151	NADH-quinone oxidoreductase subunit G	Energy production and conversion
3518552	Rv3151	NADH-quinone oxidoreductase subunit G	Energy production and conversion
3521314	Rv3154	NADH dehydrogenase I subunit J	Energy production and conversion
3522392	Rv3156	NADH-quinone oxidoreductase subunit L	Energy production and conversion
3547925	Rv3179	ATPase AAA+ superfamily protein	General function prediction only
3548638	Rv3179	ATPase AAA+ superfamily protein	General function prediction only
3556272	Rv3190c	hypothetical protein	#N/A
3570469	Rv3198c	DNA helicase II/ATP-dependent DNA helicase PcrA	Replication, recombination and repair
3581410	Rv3204	methyltransferase	Replication, recombination and repair
3588142	Rv3211	ATP-dependent RNA helicase rhIE	Replication, recombination and repair
3591060	Rv3213c	chromosome partitioning protein	Cell cycle control, cell division, chromosome partitioning
3594121	Rv3217c	hypothetical protein	#N/A
3604818	Rv3228	ribosome small subunit-dependent GTPase A	General function prediction only
3610438	Rv3234c	acyltransferase, WS/DGAT/MGAT	#N/A
3614979	Rv3239c	drug:H <sup>+</sup> antiporter-2 (14 Spanner) (DHA2) family drug resistance MFS transporter	General function prediction only
3619163	Rv3240c	protein translocase subunit secA 1	Intracellular trafficking, secretion, and vesicular transport
3622438	Rv3243c	hypothetical protein	#N/A
3625062	Rv3245c	sensor histidine kinase mtrB	Signal transduction mechanisms
3639532	Rv3259	hypothetical protein	#N/A
3654203	Rv3272	hypothetical protein	Energy production and conversion
3671862	Rv3291c	lrp/AsnC family transcriptional regulator, leucine-responsive regulatory protein	Transcription
3689520	Rv3303c	NAD(P)H dehydrogenase (quinone)	Energy production and conversion
3702370	Rv3314c	thymidine phosphorylase	Nucleotide transport and metabolism
3704767	Rv3317	succinate dehydrogenase hydrophobic membrane anchor protein	Energy production and conversion
3710553	Rv3324A	4a-hydroxytetrahydrobiopterin dehydratase	#N/A
3714382	Rv3328c	sigma-70 family RNA polymerase sigma factor	Transcription
3714928	Rv3329	hypothetical protein	#N/A

3718356	Rv3331	sugar porter (SP) family MFS transporter	Carbohydrate transport and metabolism
3721977	Rv3335c	hypothetical protein	Function unknown
3731117	Rv3343c	PPE family protein	Cell motility
3731632	Rv3343c	PPE family protein	Cell motility
3736799	Rv3343c	PPE family protein	Cell motility
3738702	Rv3345c	hypothetical protein	#N/A
3738706	Rv3345c	hypothetical protein	#N/A
3738711	Rv3345c	hypothetical protein	#N/A
3738823	Rv3345c	hypothetical protein	#N/A
3738891	Rv3345c	hypothetical protein	#N/A
3746333	Rv3347c	hypothetical protein	Cell motility
3746580	Rv3347c	hypothetical protein	Cell motility
3748367	Rv3347c	hypothetical protein	Cell motility
3776877	Rv3365c	hypothetical protein	Signal transduction mechanisms
3776919	Rv3365c	hypothetical protein	Signal transduction mechanisms
3798275	Rv3383c	geranylgeranyl diphosphate synthase, type I intergenic space or other non-protein-coding region	Coenzyme transport and metabolism
3799415	IGS		#N/A
3803289	Rv3388	PE-PGRS family protein	Function unknown
3812911	Rv3396c	GMP synthase	Nucleotide transport and metabolism
3814982	Rv3397c	squalene synthase	Lipid transport and metabolism
3817297	Rv3399	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
3823339	Rv3403c	hypothetical protein	Function unknown
3825740	Rv3406	taurine dioxygenase	Secondary metabolites biosynthesis, transport, and catabolism
3826864	Rv3408	toxin	General function prediction only
3829950	Rv3410c	IMP dehydrogenase	Nucleotide transport and metabolism
3839051	Rv3420c	ribosomal-protein-alanine acetyltransferase	General function prediction only
3859556	Rv3439c	hypothetical protein	RNA processing and modification
3860073	Rv3440c	hypothetical protein	#N/A
3860468	Rv3441c	phosphoglucosamine mutase	Carbohydrate transport and metabolism
3865174	Rv3447c	ESX-4 secretion system protein eccC4	Cell cycle control, cell division, chromosome partitioning
3873571	Rv3451	cutinase	#N/A
3877600	Rv3456c	50S ribosomal protein L17	Translation, ribosomal structure and biogenesis
3885050	Rv3467	hypothetical protein	#N/A
3885085	Rv3467	hypothetical protein	#N/A
3886065	Rv3468c	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
3886134	Rv3468c	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
3889056	Rv3471c	hypothetical protein	Carbohydrate transport and metabolism
3892850	Rv3476c	MFS transporter, MHS family, alpha- ketoglutarate permease	Carbohydrate transport and metabolism

3898587	Rv3479	hypothetical protein	General function prediction only
3927565	Rv3507	PE-PGRS family protein	Function unknown
3927596	Rv3507	PE-PGRS family protein	Function unknown
3933142	Rv3508	PE-PGRS family protein	Function unknown
3942660	IGS	intergenic space or other non-protein-coding region	#N/A
3942819	IGS	intergenic space or other non-protein-coding region	#N/A
3943198	IGS	intergenic space or other non-protein-coding region	#N/A
3943313	IGS	intergenic space or other non-protein-coding region	#N/A
3944620	IGS	intergenic space or other non-protein-coding region	#N/A
3944627	IGS	intergenic space or other non-protein-coding region	#N/A
3944667	IGS	intergenic space or other non-protein-coding region	#N/A
3952979	Rv3516	enoyl-CoA hydratase	Lipid transport and metabolism
3954422	Rv3517	hypothetical protein	#N/A
3958582	Rv3521	hypothetical protein	General function prediction only
3959597	Rv3522	hypothetical protein	Lipid transport and metabolism
3975501	IGS	intergenic space or other non-protein-coding region	#N/A
3983450	Rv3544c	hypothetical protein	Lipid transport and metabolism
3987229	Rv3547	deazaflavin-dependent nitroreductase	#N/A
4005786	Rv3564	hypothetical protein	Lipid transport and metabolism
4014151	Rv3572	hypothetical protein	#N/A
4024452	Rv3581c	2-C-methyl-D-erythritol 2,4-cyclo diphosphate synthase	Lipid transport and metabolism
4027078	Rv3585	DNA repair protein RadA	Post-translational modification, protein turnover, and chaperones
4034276	Rv3592	hypothetical protein	General function prediction only
4035006	Rv3593	hypothetical protein	Defense mechanisms
4037462	Rv3595c	PE-PGRS family protein	#N/A
4038466	Rv3596c	ATP-dependent Clp protease ATP-binding subunit ClpC	Post-translational modification, protein turnover, and chaperones
4053229	Rv3611	hypothetical protein	#N/A
4055980	Rv3616c	ESX-1 secretion-associated protein EspA	#N/A
4060083	IGS	intergenic space or other non-protein-coding region	#N/A
4069471	Rv3630	hypothetical protein	General function prediction only
4076136	IGS	intergenic space or other non-protein-coding region	#N/A
4078235	Rv3639c	hypothetical protein	#N/A
4078559	Rv3639c	hypothetical protein	#N/A
4088525	Rv3648c	CspA family cold shock protein	Transcription
4093526	IGS	intergenic space or other non-protein-coding region	#N/A
4101153	IGS	intergenic space or other non-protein-coding region	#N/A
4111481	Rv3669	hypothetical protein	#N/A
4130099	Rv3688c	hypothetical protein	Function unknown
4135290	Rv3693	hypothetical protein	General function prediction only

4143055	Rv3700c	hypothetical protein	Amino acid transport and metabolism
4148850	Rv3705c	hypothetical protein	#N/A
4156277	Rv3711c	exonuclease, DNA polymerase III, epsilon subunit	Replication, recombination and repair
4162517	Rv3719	hypothetical protein	Energy production and conversion
4164122	Rv3720	cyclopropane-fatty-acyl-phospholipid synthase	#N/A
4168461	IGS	intergenic space or other non-protein-coding region	#N/A
4168768	Rv3723	hypothetical protein	Inorganic ion transport and metabolism
4174742	Rv3727	oxidoreductase	Function unknown
4182873	Rv3731	DNA ligase (ATP)	Replication, recombination and repair
4187663	Rv3736	hypothetical protein	Transcription
4187995	Rv3737	hypothetical protein	Function unknown
4204619	Rv3759c	osmoprotectant transporter substrate-binding protein	Cell wall/membrane/envelope biogenesis
4210452	Rv3764c	two-component system, OmpR family, sensor kinase	Signal transduction mechanisms
4220352	Rv3775	hypothetical protein	Defense mechanisms
4222251	IGS	intergenic space or other non-protein-coding region	#N/A
4223060	Rv3777	oxidoreductase	Energy production and conversion
4233477	Rv3786c	hypothetical protein	Cell wall/membrane/envelope biogenesis
4242821	Rv3793	arabinosyltransferase C	#N/A
4256100	Rv3799c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
4257398	Rv3800c	polyketide synthase 13	Secondary metabolites biosynthesis, transport, and catabolism
4264396	Rv3802c	hypothetical protein	#N/A
4290677	Rv3823c	membrane transporter mmpL8	General function prediction only
4291146	Rv3823c	membrane transporter mmpL8	General function prediction only
4293250	Rv3824c	SL659 acyltransferase papA1	Secondary metabolites biosynthesis, transport, and catabolism
4302214	Rv3827c	transposase	Replication, recombination and repair
4306333	Rv3831	hypothetical protein	#N/A
4307357	Rv3833	AraC family transcriptional regulator	Transcription
4311439	Rv3837c	phosphoglycerate mutase	Carbohydrate transport and metabolism
4315562	Rv3842c	glycerophosphoryl diester phosphodiesterase	Energy production and conversion
4338909	IGS	intergenic space or other non-protein-coding region	#N/A
4340507	Rv3864	hypothetical protein	#N/A
4340519	Rv3864	hypothetical protein	#N/A
4351216	Rv3872	PE family protein	#N/A
4356287	Rv3877	ESX-1 secretion system protein eccD1	#N/A
4360729	Rv3881c	ESX-1 secretion-associated protein EspB	Inorganic ion transport and metabolism
4366449	Rv3884c	ESX-2 secretion system protein EccA2	Post-translational modification, protein turnover, and chaperones
4373673	Rv3889c	ESX-2 secretion-associated protein EspG2	#N/A

4374304	Rv3891c	esat-6 like protein EsxD	#N/A
4375805	Rv3892c	PPE family protein	#N/A
4378268	Rv3894c	type VII secretion protein EccCa	Cell cycle control, cell division, chromosome partitioning
4379857	Rv3894c	type VII secretion protein EccCa	Cell cycle control, cell division, chromosome partitioning
4382231	Rv3896c	hypothetical protein	General function prediction only
4382452	Rv3896c	hypothetical protein	General function prediction only
4391730	Rv3906c	hypothetical protein	#N/A
4401683	Rv3912	hypothetical protein	#N/A
4408022	Rv3919c	ribosomal RNA small subunit methyltransferase G	#N/A
4408332	Rv3919c	ribosomal RNA small subunit methyltransferase G	#N/A
4408726	Rv3920c	spolIJ-associated protein	General function prediction only

Tabla Anexo 6. Variantes genéticas de *Mycobacterium tuberculosis* INS\_MDR

Posicion	Gene_ID	Producto	COG
1977	IGS	intergenic space or other non-protein-coding region	#N/A
4013	RV0003	DNA replication and repair protein RecF	Replication, recombination and repair
7362	RV0006	DNA gyrase subunit A	Replication, recombination and repair
7585	RV0006	DNA gyrase subunit A	Replication, recombination and repair
8040	RV0006	DNA gyrase subunit A	Replication, recombination and repair
9304	RV0006	DNA gyrase subunit A	Replication, recombination and repair
11879	RV0008c	membrane protein	#N/A
14251	RV0012	hypothetical protein	Function unknown
17608	RV0015c	serine/threonine protein kinase	General function prediction only
21795	RV0018c	PP2C-family Ser/Thr phosphatase	Signal transduction mechanisms
26959	IGS	intergenic space or other non-protein-coding region	#N/A
27463	IGS	intergenic space or other non-protein-coding region	#N/A
32411	RV0029	hypothetical protein	#N/A
33457	RV0030	hypothetical protein	#N/A
33551	RV0030	hypothetical protein	#N/A
34044	IGS	intergenic space or other non-protein-coding region	#N/A
37031	RV0034	hypothetical protein	General function prediction only
37763	RV0035	fatty-acid-CoA ligase FadD34	Lipid transport and metabolism
42967	RV0040c	secreted proline rich protein mtc28	#N/A
55208	RV0050	peptidoglycan glycosyltransferase	Cell wall/membrane/envelope biogenesis
55250	RV0050	peptidoglycan glycosyltransferase	Cell wall/membrane/envelope biogenesis
60063	RV0057	hypothetical protein	#N/A
62046	RV0058	replicative DNA helicase	Replication, recombination and repair
63768	RV0059	hypothetical protein	#N/A
66357	RV0062	endoglucanase	Carbohydrate transport and metabolism

70813	RV0064	hypothetical protein	Function unknown
71333	RV0064	hypothetical protein	Function unknown
74056	RV0066c	isocitrate dehydrogenase	Energy production and conversion
75937	RV0068	oxidoreductase	Lipid transport and metabolism
80613	IGS	intergenic space or other non-protein-coding region	#N/A
92196	RV0083	oxidoreductase	Energy production and conversion
93141	RV0084	formate hydrogenlyase hycD	Energy production and conversion
105042	RV0095c	hypothetical protein	#N/A
122106	RV0103c	cation-transporting P-type ATPase B	Inorganic ion transport and metabolism
128001	RV0107c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
130224	RV0107c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
133837	IGS	intergenic space or other non-protein-coding region	#N/A
133860	IGS	intergenic space or other non-protein-coding region	#N/A
146085	RV0120c	elongation factor G-like protein	Translation, ribosomal structure and biogenesis
149321	IGS	intergenic space or other non-protein-coding region	#N/A
154189	IGS	intergenic space or other non-protein-coding region	#N/A
154281	RV0127	hypothetical protein	Carbohydrate transport and metabolism
157290	RV0129c	antigen 85-C	General function prediction only
161085	RV0133	acetyltransferase	General function prediction only
162579	RV0134	epoxide hydrolase EphF	General function prediction only
177855	RV0151c	PE family protein	#N/A
178939	RV0151c	PE family protein	#N/A
180023	RV0152c	PE family protein	#N/A
182220	RV0154c	acyl-CoA dehydrogenase	Lipid transport and metabolism
184459	RV0155	NAD(P)(+) transhydrogenase (AB-specific), alpha subunit	Energy production and conversion
184762	RV0156	NAD(P) transhydrogenase subunit alpha	Energy production and conversion
188798	RV0159c	PE family protein	#N/A
194679	RV0165c	GntR family transcriptional regulator	#N/A



196520	RV0166	fatty-acyl-CoA synthase	Lipid transport and metabolism
196640	RV0166	fatty-acyl-CoA synthase	Lipid transport and metabolism
200152	RV0170	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
203267	RV0172	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
206337	RV0174	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
207224	RV0175	MCE-associated membrane protein	#N/A
212351	RV0181c	hypothetical protein	General function prediction only
223940	RV0192	hypothetical protein	Function unknown
225321	RV0193c	hypothetical protein	#N/A
227096	RV0194	ATP-binding cassette, subfamily C	Defense mechanisms
231112	RV0195	two component system transcriptional regulator, luxR-family	Signal transduction mechanisms
234448	RV0197	nitrate reductase catalytic subunit	Energy production and conversion
261869	RV0218	hypothetical protein	General function prediction only
265554	RV0222	enoyl-CoA hydratase	Lipid transport and metabolism
278681	RV0233	R2-like ligand binding oxidase	Nucleotide transport and metabolism
283614	RV0236c	hypothetical protein	#N/A
285772	RV0236c	hypothetical protein	#N/A
285871	RV0236c	hypothetical protein	#N/A
294586	RV0244c	acyl-CoA dehydrogenase fadE5	Lipid transport and metabolism
305362	RV0252	nitrite reductase [NAD(P)H], large subunit	Energy production and conversion
310973	RV0259c	hypothetical protein	Function unknown
311613	RV0260c	uroporphyrinogen-III synthase	Coenzyme transport and metabolism
316129	RV0264c	hypothetical protein	Amino acid transport and metabolism
327897	RV0271c	acyl-CoA dehydrogenase fadE6	Lipid transport and metabolism
336692	RV0279c	WAG22 antigen	#N/A
337820	RV0279c	WAG22 antigen	#N/A
340372	RV0280	PPE family protein	Cell motility
346275	RV0284	ESX-3 secretion system protein eccC3	Cell cycle control, cell division, chromosome partitioning

356528	RV0292	ESX-3 secretion system protein EccE3	#N/A
376774	RV0307c	hypothetical protein	#N/A
384380	RV0315	hypothetical protein	Carbohydrate transport and metabolism
386432	RV0318c	ZIP family zinc transporter	Inorganic ion transport and metabolism
387353	RV0319	pyrrolidone-carboxylate peptidase	Post-translational modification, protein turnover, and chaperones
398116	RV0332	hypothetical protein	#N/A
403364	RV0338c	oxidoreductase	Energy production and conversion
403920	RV0338c	oxidoreductase	Energy production and conversion
403980	RV0338c	oxidoreductase	Energy production and conversion
404326	RV0338c	oxidoreductase	Energy production and conversion
408723	RV0340	hypothetical protein	#N/A
414486	RV0344c	lipoprotein LpqJ	Lipid transport and metabolism
420008	RV0350	chaperone dnaK	Post-translational modification, protein turnover, and chaperones
425343	RV0355c	PPE family protein	#N/A
440561	RV0362	magnesium transporter	Inorganic ion transport and metabolism
444352	RV0366c	hypothetical protein	Function unknown
457453	RV0381c	hypothetical protein	#N/A
463337	IGS	intergenic space or other non-protein-coding region	#N/A
472708	IGS	intergenic space or other non-protein-coding region	#N/A
475181	RV0395	hypothetical protein	#N/A
479635	RV0400c	glutaryl-CoA dehydrogenase	Lipid transport and metabolism
489938	RV0405/RV0406c	polyketide synthase 12//beta lactamase like protein	#N/A
498560	RV0412c	hypothetical protein	#N/A
501538	RV0415	glycine oxidase ThiO	Amino acid transport and metabolism
502592	RV0417	thiazole synthase	Coenzyme transport and metabolism
503357	IGS	intergenic space or other non-protein-coding region	#N/A
513260	RV0425c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
534694	RV0446c	hypothetical protein	Function unknown

535698	RV0447c	cyclopropane-fatty-acyl-phospholipid synthase	Cell wall/membrane/envelope biogenesis
541204	RV0450c	transmembrane transporter mmpL4	General function prediction only
551528	RV0459	hypothetical protein	Function unknown
558504	RV0467	isocitrate lyase	Energy production and conversion
563423	RV0472c	TetR family transcriptional regulator	Transcription
573265	RV0484c	oxidoreductase	General function prediction only
584441	RV0493c	hypothetical protein	#N/A
586217	RV0495c	hypothetical protein	#N/A
589539	RV0499	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
590439	RV0500	pyrroline-5-carboxylate reductase	Amino acid transport and metabolism
597819	RV0507	transmembrane transporter mmpL2	General function prediction only
598478	RV0507	transmembrane transporter mmpL2	General function prediction only
610123	IGS	intergenic space or other non-protein-coding region	#N/A
630725	RV0538	hypothetical protein	General function prediction only
635636	RV0543c	hypothetical protein	#N/A
637322	RV0545c	PiT family inorganic phosphate transporter	Inorganic ion transport and metabolism
637925	RV0546c	lactoylglutathione lyase	General function prediction only
648005	RV0556	hypothetical protein	#N/A
664252	RV0571c	hypothetical protein	General function prediction only
665296	RV0572c	hypothetical protein	#N/A
669401	RV0575c	oxidoreductase	Coenzyme transport and metabolism
673241	RV0578c	PE-PGRS family protein	#N/A
675676	RV0578c	PE-PGRS family protein	#N/A
675800	RV0578c	PE-PGRS family protein	#N/A
685464	RV0587	ABC transporter permease	Secondary metabolites biosynthesis, transport, and catabolism
685611	RV0587	ABC transporter permease	Secondary metabolites biosynthesis, transport, and catabolism
686975	RV0589	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism

690468	RV0591	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
697117	RV0597A	hypothetical protein	#N/A
698971	RV0601c	sensor histidine kinase component HK2	Signal transduction mechanisms
709229	RV0613c	hypothetical protein	#N/A
726706	RV0631c	exodeoxyribonuclease V subunit gamma	Replication, recombination and repair
736713	RV0642c	hydroxymycolate synthase MmaA4	Cell wall/membrane/envelope biogenesis
752346	RV0655	ABC transporter ATP-binding protein	Secondary metabolites biosynthesis, transport, and catabolism
752805	IGS	intergenic space or other non-protein-coding region	#N/A
754189	RV0658c	hypothetical protein	General function prediction only
761113	RV0667	DNA-directed RNA polymerase subunit beta	Transcription
764998	RV0668	DNA-directed RNA polymerase subunit beta'	Transcription
769666	IGS	intergenic space or other non-protein-coding region	#N/A
773812	RV0673	enoyl-CoA hydratase	Lipid transport and metabolism
775642	RV0676c	transmembrane transporter mmpL5	General function prediction only
781398	IGS	intergenic space or other non-protein-coding region	#N/A
796512	RV0696	mycofactocin system glycosyltransferase	Cell wall/membrane/envelope biogenesis
812811	RV0717	30S ribosomal protein S14	Translation, ribosomal structure and biogenesis
820486	RV0727c	L-ribulose-5-phosphate 4-epimerase	Carbohydrate transport and metabolism
826094	RV0732	preprotein translocase subunit secY	Intracellular trafficking, secretion, and vesicular transport
839952	RV0747	PE-PGRS family protein	#N/A
841767	RV0749A	hypothetical protein	#N/A
852913	RV0758	two-component system, OmpR family, sensor kinase	Signal transduction mechanisms
857699	RV0764c	lanosterol 14-alpha demethylase	Secondary metabolites biosynthesis, transport, and catabolism
862990	RV0769	3-oxoacyl-ACP reductase	Lipid transport and metabolism

863808	RV0770	dehydrogenase/reductase	Lipid transport and metabolism
880565	RV0785	KsdD-like steroid dehydrogenase	General function prediction only
882260	RV0787	hypothetical protein	#N/A
885545	RV0791c	hypothetical protein	Energy production and conversion
893736	RV0800	aminopeptidase	Amino acid transport and metabolism
900224	RV0806c	exopolysaccharide phosphotransferase CpsY	#N/A
903553	RV0808	amidophosphoribosyltransferase	Nucleotide transport and metabolism
903916	RV0809	phosphoribosylformylglycinamide cyclo-ligase	Nucleotide transport and metabolism
906860	RV0812	4-amino-4-deoxychorismate lyase	Amino acid transport and metabolism
907819	RV0813c	hypothetical protein	#N/A
919396	RV0825c	hypothetical protein	Transcription
927388	RV0833	PE-PGRS family protein	Function unknown
942482	IGS	intergenic space or other non-protein-coding region	#N/A
945217	RV0848	cysteine synthase A	Amino acid transport and metabolism
947432	RV0850	transposase	#N/A
949538	RV0853c	alpha-keto-acid decarboxylase	Carbohydrate transport and metabolism
955014	IGS	intergenic space or other non-protein-coding region	#N/A
955527	RV0859	acetyl-CoA C-acetyltransferase	Lipid transport and metabolism
958925	RV0861c	DNA excision repair protein ERCC-3	#N/A
979707	RV0881	methyltransferase	Translation, ribosomal structure and biogenesis
986466	IGS	intergenic space or other non-protein-coding region	#N/A
989722	RV0889c	citrate synthase	Energy production and conversion
991518	RV0890c	LuxR family transcriptional regulator	General function prediction only
993349	RV0891c	transcriptional regulator	Signal transduction mechanisms
996200	RV0893c	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
997902	RV0895	acyltransferase, WS/DGAT/MGAT	#N/A
1029253	RV0922	transposase	Replication, recombination and repair
1037016	RV0930	phosphate transporter permease pstA 1	Inorganic ion transport and metabolism

1040054	RV0932c	phosphate-binding protein pstS 2	Inorganic ion transport and metabolism
1042229	RV0934	phosphate-binding protein pstS 1	Inorganic ion transport and metabolism
1047169	RV0938	DNA ligase D	Replication, recombination and repair
1054366	RV0945	oxidoreductase	General function prediction only
1057792	RV0948c	chorismate mutase	Amino acid transport and metabolism
1064520	RV0953c	oxidoreductase	Energy production and conversion
1068155	RV0956	phosphoribosylglycinamide formyltransferase	Nucleotide transport and metabolism
1068436	RV0957	bifunctional purine biosynthesis protein purH	Nucleotide transport and metabolism
1069454	RV0957	bifunctional purine biosynthesis protein purH	Nucleotide transport and metabolism
1070706	RV0958	magnesium chelatase subunit I	Coenzyme transport and metabolism
1073037	RV0959	hypothetical protein	General function prediction only
1074562	RV0962c	lipoprotein LprP	#N/A
1075283	IGS	intergenic space or other non-protein-coding region	#N/A
1076313	RV0964c	hypothetical protein	#N/A
1077316	RV0966c	hypothetical protein	#N/A
1079931	RV0969	heavy metal translocating P-type ATPase	#N/A
1081685	RV0970	hypothetical protein	#N/A
1087197	RV0974c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
1087283	RV0974c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
1091815	RV0977	PE-PGRS family protein	Function unknown
1093326	IGS	intergenic space or other non-protein-coding region	#N/A
1096636	IGS	intergenic space or other non-protein-coding region	#N/A
1100238	RV0983	serine protease PepD	Post-translational modification, protein turnover, and chaperones
1103253	RV0987	ABC transporter permease	Cell wall/membrane/envelope biogenesis
1106426	RV0989c	heptaprenyl diphosphate synthase	Coenzyme transport and metabolism
1107438	IGS	intergenic space or other non-protein-coding region	#N/A
1107921	RV0990c	hypothetical protein	Cell motility

1109979	RV0993	UTP-glucose-1-phosphate uridylyltransferase	Cell wall/membrane/envelope biogenesis
1126893	RV1007c	methionyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1127652	RV1008	TatD family hydrolase	Replication, recombination and repair
1142270	RV1020	transcription-repair-coupling factor	Replication, recombination and repair
1144668	RV1023	enolase	Carbohydrate transport and metabolism
1149555	RV1028c	sensor protein kdpD	Signal transduction mechanisms
1150589	RV1028c	sensor protein kdpD	Signal transduction mechanisms
1162278	RV1039c	PPE family protein	Cell motility
1163138	RV1040c	PE family protein	#N/A
1164575	IGS	intergenic space or other non-protein- coding region	#N/A
1167651	RV1044	hypothetical protein	Transcription
1177999	RV1056	hypothetical protein	Function unknown
1178121	RV1056	hypothetical protein	Function unknown
1190159	RV1067c	PE-PGRS family protein	Function unknown
1199552	RV1075c	hypothetical protein	Amino acid transport and metabolism
1200423	IGS	intergenic space or other non-protein- coding region	#N/A
1212081	RV1087	PE-PGRS family protein	Function unknown
1212331	RV1087	PE-PGRS family protein	Function unknown
1212340	RV1087	PE-PGRS family protein	Function unknown
1218306	RV1091	PE-PGRS family protein	#N/A
1218742	RV1091	PE-PGRS family protein	#N/A
1220685	RV1093	serine hydroxymethyltransferase 1	#N/A
1224372	IGS	intergenic space or other non-protein- coding region	#N/A
1248983	RV1125	hypothetical protein	#N/A
1272015	IGS	intergenic space or other non-protein- coding region	#N/A
1281123	RV1154c	hypothetical protein	Function unknown
1292107	RV1162	nitrate reductase subunit beta	Energy production and conversion
1307603	RV1175c	2,4-dienoyl-CoA reductase	Energy production and conversion
1324730	RV1184c	hypothetical protein	#N/A
1327896	RV1186c	hypothetical protein	Signal transduction mechanisms
1328693	RV1186c	hypothetical protein	Signal transduction mechanisms

1349007	RV1205	hypothetical protein	General function prediction only
1354443	RV1211	hypothetical protein	#N/A
1360215	RV1217c	antibiotic transporter permease	Cell wall/membrane/envelope biogenesis
1363365	RV1219c	transcriptional regulator	Transcription
1372893	RV1229c	ATP-binding protein	Cell cycle control, cell division, chromosome partitioning
1374071	RV1230c	membrane protein	Cell wall/membrane/envelope biogenesis
1375730	RV1232c	hypothetical protein	Inorganic ion transport and metabolism
1382634	RV1239c	magnesium and cobalt transporter CorA	Inorganic ion transport and metabolism
1389744	RV1248c	2-oxoglutarate decarboxylase	Energy production and conversion
1393632	RV1249c	membrane protein	#N/A
1396928	RV1251c	hypothetical protein	General function prediction only
1404175	RV1256c	cytochrome P450 130 cyp130	Secondary metabolites biosynthesis, transport, and catabolism
1411216	RV1263	amidase	Translation, ribosomal structure and biogenesis
1413154	IGS	intergenic space or other non-protein-coding region	#N/A
1428512	RV1278	hypothetical protein	Function unknown
1431811	RV1280c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
1440475	RV1286	bifunctional enzyme CysN/CysC	Inorganic ion transport and metabolism
1445480	IGS	intergenic space or other non-protein-coding region	#N/A
1445787	RV1291c	hypothetical protein	#N/A
1446929	RV1292	arginyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1452077	RV1296	homoserine kinase	Amino acid transport and metabolism
1457150	RV1300	HemK family methyltransferase	Translation, ribosomal structure and biogenesis
1480030	RV1318c	adenylate cyclase	Signal transduction mechanisms
1480978	RV1319c	adenylate cyclase	Signal transduction mechanisms
1482633	RV1320c	adenylate cyclase	Signal transduction mechanisms



1484714	RV1321	hypothetical protein	Replication, recombination and repair
1499280	RV1330c	nicotinate phosphoribosyltransferase pncB1	Coenzyme transport and metabolism
1512037	RV1347c	lysine N-acyltransferase mbtK	Translation, ribosomal structure and biogenesis
1519853	RV1353c	transcriptional regulator	Transcription
1523309	RV1355c	molybdopterin biosynthesis protein MoeY	Coenzyme transport and metabolism
1526825	RV1358	transcriptional regulator	General function prediction only
1532302	RV1360	oxidoreductase	Energy production and conversion
1535147	RV1363c	membrane protein	#N/A
1536257	RV1364c	anti-anti-sigma factor	Signal transduction mechanisms
1546536	RV1373	glycolipid sulfotransferase	#N/A
1547131	RV1374c	hypothetical protein	#N/A
1552553	RV1378c	hypothetical protein	#N/A
1563723	RV1388	integration host factor mihF	#N/A
1570572	RV1394c	cytochrome P450 132 cyp132	Secondary metabolites biosynthesis, transport, and catabolism
1573021	RV1396c	PE-PGRS family protein	Cell wall/membrane/envelope biogenesis
1573666	RV1396c	PE-PGRS family protein	Cell wall/membrane/envelope biogenesis
1575201	RV1399c	lipase lipH	Lipid transport and metabolism
1586255	RV1410c	MFS-type drug efflux transporter P55	Carbohydrate transport and metabolism
1588905	RV1412	riboflavin synthase subunit alpha	Coenzyme transport and metabolism
1609846	RV1431	hypothetical protein	#N/A
1612295	RV1434	hypothetical protein	#N/A
1613041	IGS	intergenic space or other non-protein-coding region	#N/A
1613966	RV1436	glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate transport and metabolism
1618637	RV1441c	triacylglycerol lipase	#N/A
1618639	RV1441c	triacylglycerol lipase	#N/A
1618649	RV1441c	triacylglycerol lipase	#N/A
1618652	RV1441c	triacylglycerol lipase	#N/A
1618678	RV1441c	triacylglycerol lipase	#N/A
1618686	RV1441c	triacylglycerol lipase	#N/A
1624797	RV1446c	opcA protein	Carbohydrate transport and metabolism

1628595	RV1449c	transketolase	Carbohydrate transport and metabolism
1630154	RV1449c	transketolase	Carbohydrate transport and metabolism
1636832	RV1452c	PE-PGRS family protein	#N/A
1636924	RV1452c	PE-PGRS family protein	#N/A
1636986	RV1452c	PE-PGRS family protein	#N/A
1639600	RV1453	hypothetical protein	Signal transduction mechanisms
1645808	RV1459c	hypothetical protein	#N/A
1650078	RV1462	Fe-S cluster assembly protein SufD	Post-translational modification, protein turnover, and chaperones
1651312	RV1463	FeS assembly ATPase SufC	Post-translational modification, protein turnover, and chaperones
1676296	RV1486c	hypothetical protein	#N/A
1689355	RV1498c	methyltransferase	Coenzyme transport and metabolism
1692147	RV1501	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
1692801	IGS	intergenic space or other non-protein-coding region	#N/A
1693567	RV1502	hypothetical protein	#N/A
1698917	RV1508c	membrane protein	#N/A
1706125	RV1514c	hypothetical protein	Cell wall/membrane/envelope biogenesis
1711633	RV1520	sugar transferase	#N/A
1717687	RV1523	methyltransferase	Coenzyme transport and metabolism
1718767	RV1524	glycosyltransferase	Carbohydrate transport and metabolism
1728843	IGS	intergenic space or other non-protein-coding region	#N/A
1732245	RV1530	alcohol dehydrogenase	Amino acid transport and metabolism
1735000	RV1534	transcriptional regulator	Transcription
1740777	RV1537	DNA polymerase IV 1	#N/A
1752567	RV1548c	PPE family protein	Cell motility
1755525	RV1551	glycerol-3-phosphate O-acyltransferase	Lipid transport and metabolism
1759258	RV1552	fumarate reductase flavoprotein subunit	Energy production and conversion
1760298	RV1554	fumarate reductase subunit C	Energy production and conversion
1762621	RV1557	hypothetical protein	General function prediction only

1773693	RV1565c	hypothetical protein	Lipid transport and metabolism
1775318	IGS	intergenic space or other non-protein-coding region	#N/A
1777188	RV1569	8-amino-7-oxononanoate synthase	Coenzyme transport and metabolism
1777219	RV1569	8-amino-7-oxononanoate synthase	Coenzyme transport and metabolism
1778436	RV1570	dethiobiotin synthetase	Coenzyme transport and metabolism
1779249	RV1572c	hypothetical protein	#N/A
1782625	IGS	intergenic space or other non-protein-coding region	#N/A
1782717	IGS	intergenic space or other non-protein-coding region	#N/A
1782774	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782780	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782791	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782793	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782799	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1793776	IGS	intergenic space or other non-protein-coding region	#N/A
1797034	RV1595	L-aspartate oxidase	Coenzyme transport and metabolism
1798362	RV1597	hypothetical protein	Coenzyme transport and metabolism
1803272	RV1602	imidazole glycerol phosphate synthase subunit HisH	Amino acid transport and metabolism
1804416	RV1604	myo-inositol-1(or 4)-monophosphatase	Carbohydrate transport and metabolism
1805955	RV1606	phosphoribosyl-AMP cyclohydrolase	Amino acid transport and metabolism
1815611	RV1615	hypothetical protein	Function unknown
1817983	RV1618	acyl-CoA thioesterase II	Lipid transport and metabolism
1819598	RV1619	hypothetical protein	Function unknown
1836293	IGS	intergenic space or other non-protein-coding region	#N/A
1847926	RV1639c	hypothetical protein	General function prediction only
1854307	RV1644	TrmH family RNA methyltransferase	Translation, ribosomal structure and biogenesis
1856784	RV1647	adenylate cyclase	Signal transduction mechanisms

1885779	RV1662	polyketide synthase 8	Secondary metabolites biosynthesis, transport, and catabolism
1895181	RV1668c	hypothetical protein	General function prediction only
1895524	IGS	intergenic space or other non-protein-coding region	#N/A
1899358	RV1674c	transcriptional regulator	Inorganic ion transport and metabolism
1901500	RV1676	hypothetical protein	Post-translational modification, protein turnover, and chaperones
1907303	RV1682	hypothetical protein	#N/A
1917979	RV1694	hemolysin TlyA family protein	Translation, ribosomal structure and biogenesis
1931186	RV1704c	AAT family amino acid transporter	Amino acid transport and metabolism
1933995	RV1706Xc	hypothetical protein	#N/A
1938034	RV1709	segregation and condensation protein A	Function unknown
1943046	RV1715	3-hydroxybutyryl-CoA dehydrogenase	Lipid transport and metabolism
1943599	RV1716	arylformamidase	General function prediction only
1944114	RV1716	arylformamidase	General function prediction only
1944409	RV1716	arylformamidase	General function prediction only
1945135	RV1718	hypothetical protein	Function unknown
1947910	RV1722	carboxylase	Lipid transport and metabolism
1950774	RV1724c	hypothetical protein	#N/A
1960291	RV1733c	hypothetical protein	#N/A
1967244	RV1739c	sulfate permease	Inorganic ion transport and metabolism
1981063	IGS	intergenic space or other non-protein-coding region	#N/A
1982908	RV1753c	PPE family protein	Cell motility
1983320	RV1753c	PPE family protein	Cell motility
1989086	RV1758	esterase/lipase	#N/A
1990926	RV1759c	WAG22 antigen	#N/A
1993815	RV1760	acyltransferase, WS/DGAT/MGAT	#N/A
1993937	RV1760	acyltransferase, WS/DGAT/MGAT	#N/A
2006039	RV1771	FAD-linked oxidoreductase	Energy production and conversion
2007552	RV1773c	hypothetical protein	Transcription
2022875	RV1784	ESX-5 secretion system protein eccCa5	#N/A
2024061	RV1785c	cytochrome P450 143 cyp143	Secondary metabolites biosynthesis, transport, and catabolism

2037723	RV1798	ESX-5 secretion system protein EccA5	General function prediction only
2044198	RV1802	PPE family protein	Cell motility
2045317	RV1803c	hypothetical protein	#N/A
2048986	RV1807	hypothetical protein	Cell motility
2049072	RV1807	hypothetical protein	Cell motility
2049104	RV1807	hypothetical protein	Cell motility
2052042	RV1809	hypothetical protein	Cell motility
2053418	IGS	intergenic space or other non-protein-coding region	#N/A
2053689	RV1811	Mg <sup>2+</sup> transporter-C (MgtC) family protein	Function unknown
2055278	RV1812c	NADH dehydrogenase-like protein	Energy production and conversion
2071583	RV1825	hypothetical protein	Function unknown
2074516	IGS	intergenic space or other non-protein-coding region	#N/A
2074572	IGS	intergenic space or other non-protein-coding region	#N/A
2077260	RV1832	glycine dehydrogenase	Amino acid transport and metabolism
2096193	RV1846c	transcriptional regulator Blal	Transcription
2097997	RV1850	urease subunit alpha	Amino acid transport and metabolism
2108148	RV1860	alanine and proline-rich secreted protein apa	Function unknown
2108845	IGS	intergenic space or other non-protein-coding region	#N/A
2116910	RV1867	acetyl-CoA C-acetyltransferase	Lipid transport and metabolism
2123176	IGS	intergenic space or other non-protein-coding region	#N/A
2123189	RV1873	hypothetical protein	Function unknown
2128877	RV1878	glutamine synthetase	Amino acid transport and metabolism
2128915	RV1878	glutamine synthetase	Amino acid transport and metabolism
2134222	RV1884c	hypothetical protein	#N/A
2135877	IGS	intergenic space or other non-protein-coding region	#N/A
2137549	RV1888c	transmembrane protein	#N/A
2143335	RV1895	chlorophyll synthesis pathway, bchC	Amino acid transport and metabolism
2144624	RV1897c	D-tyrosyl-tRNA(Tyr) deacylase	Translation, ribosomal structure and biogenesis
2147029	RV1900c	hypothetical protein	Signal transduction mechanisms
2153417	RV1907c	hypothetical protein	#N/A
2155175	RV1908c	catalase-peroxidase	Inorganic ion transport and metabolism

2162188	RV1916	isocitrate lyase	Energy production and conversion
2163797	RV1917c	hypothetical protein	Cell motility
2165293	RV1917c	hypothetical protein	Cell motility
2165561	RV1917c	hypothetical protein	Cell motility
2169005	RV1918c	PPE family protein	Cell motility
2171973	RV1920	membrane protein	Lipid transport and metabolism
2174223	RV1922	hypothetical protein	Defense mechanisms
2180381	RV1927	hypothetical protein	Function unknown
2207532	IGS	intergenic space or other non-protein-coding region	#N/A
2211834	RV1968	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2212881	RV1969	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2216256	RV1971	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2216451	RV1971	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2220520	RV1977	hypothetical protein	Post-translational modification, protein turnover, and chaperones
2223301	IGS	intergenic space or other non-protein-coding region	#N/A
2228975	IGS	intergenic space or other non-protein-coding region	#N/A
2231140	RV1987	chitinase	Carbohydrate transport and metabolism
2245540	RV2000	hypothetical protein	Energy production and conversion
2260533	IGS	intergenic space or other non-protein-coding region	#N/A
2264790	RV2017	transcriptional regulator	Amino acid transport and metabolism
2265067	IGS	intergenic space or other non-protein-coding region	#N/A
2267380	RV2022c	hypothetical protein	Function unknown
2269788	RV2024c	hypothetical protein	General function prediction only
2270110	RV2024c	hypothetical protein	General function prediction only
2278454	RV2030c	erythromycin esterase	General function prediction only
2283038	RV2037c	hypothetical protein	General function prediction only

2285259	RV2039c	multiple sugar transporter permease	Carbohydrate transport and metabolism
2286747	RV2041c	multiple sugar transporter substrate-binding protein	Carbohydrate transport and metabolism
2287129	RV2041c	multiple sugar transporter substrate-binding protein	Carbohydrate transport and metabolism
2288470	RV2042c	hypothetical protein	#N/A
2289221	RV2043c	nicotinamidase/pyrazinamidase	Secondary metabolites biosynthesis, transport, and catabolism
2306314	RV2048c	polyketide synthase 12	#N/A
2306506	RV2048c	polyketide synthase 12	#N/A
2306526	RV2048c	polyketide synthase 12	#N/A
2306626	RV2048c	polyketide synthase 12	#N/A
2306629	RV2048c	polyketide synthase 12	#N/A
2306638	RV2048c	polyketide synthase 12	#N/A
2306640	RV2048c	polyketide synthase 12	#N/A
2306650	RV2048c	polyketide synthase 12	#N/A
2306668	RV2048c	polyketide synthase 12	#N/A
2306674	RV2048c	polyketide synthase 12	#N/A
2306680	RV2048c	polyketide synthase 12	#N/A
2306689	RV2048c	polyketide synthase 12	#N/A
2306692	RV2048c	polyketide synthase 12	#N/A
2306698	RV2048c	polyketide synthase 12	#N/A
2306704	RV2048c	polyketide synthase 12	#N/A
2306707	RV2048c	polyketide synthase 12	#N/A
2306713	RV2048c	polyketide synthase 12	#N/A
2306715	RV2048c	polyketide synthase 12	#N/A
2310578	RV2051c	apolipoprotein N-acyltransferase	Cell wall/membrane/envelope biogenesis
2315677	RV2059	zinc/manganese transporter substrate-binding protein	Inorganic ion transport and metabolism
2329541	RV2072c	precorrin-6Y C(5,15)-methyltransferase	Coenzyme transport and metabolism
2334015	RV2077c	hypothetical protein	#N/A
2335083	RV2078	hypothetical protein	#N/A
2335502	RV2079	hypothetical protein	Cell wall/membrane/envelope biogenesis
2340629	RV2082	hypothetical protein	#N/A
2341644	RV2083	hypothetical protein	#N/A
2345045	RV2088	serine/threonine protein kinase	General function prediction only
2347450	IGS	intergenic space or other non-protein-coding region	#N/A
2348454	RV2090	5'-3' exonuclease	Replication, recombination and repair

2353421	RV2095c	protein pafC	Transcription
2360733	RV2101	helicase helZ	#N/A
2361612	RV2101	helicase helZ	#N/A
2362049	RV2101	helicase helZ	#N/A
2386397	RV2125	hypothetical protein	General function prediction only
2401878	IGS	intergenic space or other non-protein-coding region	#N/A
2401891	IGS	intergenic space or other non-protein-coding region	#N/A
2415664	RV2155c	UDP-N-acetylmuramoylalanine--D-glutamate ligase	#N/A
2424933	IGS	intergenic space or other non-protein-coding region	#N/A
2437979	RV2176	serine/threonine protein kinase	General function prediction only
2440934	RV2178c	3-deoxy-7-phosphoheptulonate synthase	Amino acid transport and metabolism
2451954	RV2189c	hypothetical protein	#N/A
2462879	RV2198c	membrane protein mmpS3	General function prediction only
2463102	IGS	intergenic space or other non-protein-coding region	#N/A
2478647	RV2213	leucyl aminopeptidase	Amino acid transport and metabolism
2480166	RV2214c	oxidoreductase	General function prediction only
2483959	RV2216	epimerase	General function prediction only
2484263	RV2216	epimerase	General function prediction only
2499734	RV2226	hypothetical protein	Function unknown
2502081	RV2228c	hypothetical protein	Carbohydrate transport and metabolism
2503633	RV2230c	hypothetical protein	Function unknown
2509148	RV2236c	cobalamin biosynthesis protein CobD	Coenzyme transport and metabolism
2509730	RV2237	hypothetical protein	Function unknown
2518927	RV2245	3-oxoacyl-ACP synthase	Lipid transport and metabolism
2521350	RV2247	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
2529690	RV2256c	hypothetical protein	#N/A
2531752	RV2258c	transcriptional regulator	Coenzyme transport and metabolism
2558426	RV2285	acyltransferase, WS/DGAT/MGAT	General function prediction only
2573766	IGS	intergenic space or other non-protein-coding region	#N/A



2586137	RV2314c	hypothetical protein	General function prediction only
2598410	RV2326c	cobalt/nickel transporter ATP-binding protein	General function prediction only
2612642	RV2337c	hypothetical protein	#N/A
2621068	RV2343c	DNA primase	Replication, recombination and repair
2642385	RV2360c	hypothetical protein	#N/A
2654381	RV2374c	heat-inducible transcription repressor hrcA	Transcription
2656235	RV2377c	protein mbtH	Function unknown
2660329	RV2379c	amino acid adenylation domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
2680668	RV2386A	hypothetical protein	#N/A
2695388	RV2398c	sulfate ABC transporter, permease CysW	Inorganic ion transport and metabolism
2713805	IGS	intergenic space or other non-protein-coding region	#N/A
2734084	RV2436	ribokinase	Carbohydrate transport and metabolism
2736444	RV2438c	glutamine-dependent NAD(+) synthetase	Coenzyme transport and metabolism
2745899	RV2446c	hypothetical protein	#N/A
2752708	IGS	intergenic space or other non-protein-coding region	#N/A
2754927	RV2455c	2-oxoacid:acceptor oxidoreductase subunit alpha	Energy production and conversion
2760162	RV2458	homocysteine S-methyltransferase	Amino acid transport and metabolism
2775052	RV2471	alpha-glucosidase	Carbohydrate transport and metabolism
2779146	RV2476c	glutamate dehydrogenase	Amino acid transport and metabolism
2783931	RV2477c	ChvD family ATP-binding cassette protein	General function prediction only
2786962	RV2482c	glycerol-3-phosphate acyltransferase	Lipid transport and metabolism
2805163	RV2490c	PE-PGRS family protein	#N/A
2805206	RV2490c	PE-PGRS family protein	#N/A
2814971	RV2501c	acetyl/propionyl-CoA carboxylase subunit alpha	Lipid transport and metabolism
2818847	RV2503c	3-oxoacid CoA-transferase subunit B	Lipid transport and metabolism
2821352	RV2505c	fatty-acyl-CoA synthase	Lipid transport and metabolism
2821622	IGS	intergenic space or other non-protein-coding region	#N/A
2827994	IGS	intergenic space or other non-protein-coding region	#N/A

2828029	IGS	intergenic space or other non-protein-coding region	#N/A
2830535	RV2513	hypothetical protein	#N/A
2855269	RV2531c	arginine decarboxylase	Amino acid transport and metabolism
2861361	RV2537c	3-dehydroquinate dehydratase	Amino acid transport and metabolism
2865770	RV2542	hypothetical protein	Replication, recombination and repair
2865892	RV2542	hypothetical protein	Replication, recombination and repair
2871058	RV2551c	leader peptidase (prepilin peptidase)/N-methyltransferase	Cell motility
2880712	RV2560	hypothetical protein	Function unknown
2881465	RV2561	hypothetical protein	#N/A
2888211	RV2566	hypothetical protein	Function unknown
2889643	RV2566	hypothetical protein	Function unknown
2891277	RV2567	hypothetical protein	Function unknown
2891738	RV2567	hypothetical protein	Function unknown
2893248	RV2568c	hypothetical protein	Function unknown
2894218	RV2569c	hypothetical protein	Amino acid transport and metabolism
2910471	RV2584c	adenine phosphoribosyltransferase	Nucleotide transport and metabolism
2911303	RV2585c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
2912304	RV2585c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
2923401	RV2592c	Holliday junction DNA helicase subunit RuvB	Replication, recombination and repair
2927949	IGS	intergenic space or other non-protein-coding region	#N/A
2939383	RV2611c	phosphatidylinositol mannoside acyltransferase	Cell wall/membrane/envelope biogenesis
2939667	RV2611c	phosphatidylinositol mannoside acyltransferase	Cell wall/membrane/envelope biogenesis
2945177	IGS	intergenic space or other non-protein-coding region	#N/A
2960241	RV2634c	PE-PGRS family protein	#N/A
2966455	IGS	intergenic space or other non-protein-coding region	#N/A
2968923	IGS	intergenic space or other non-protein-coding region	#N/A
2974943	RV2650c	hypothetical protein	General function prediction only
2980921	IGS	intergenic space or other non-protein-coding region	#N/A

2984750	RV2668	hypothetical protein	#N/A
2987651	IGS	intergenic space or other non-protein-coding region	#N/A
3004101	RV2687c	fluoroquinolones export permease	#N/A
3005195	RV2688c	fluoroquinolones export ATP-binding protein	Defense mechanisms
3009702	RV2691	trk system potassium uptake protein TrkA	Inorganic ion transport and metabolism
3017475	RV2702	polyphosphate glucokinase	Transcription
3027592	RV2714	hypothetical protein	#N/A
3029620	RV2716	phenazine biosynthesis protein PhzF family protein	General function prediction only
3031682	IGS	intergenic space or other non-protein-coding region	#N/A
3033436	RV2721c	hypothetical protein	Cell wall/membrane/envelope biogenesis
3041881	RV2729c	hypothetical protein	#N/A
3054091	RV2741	PE-PGRS family protein	#N/A
3054331	RV2741	PE-PGRS family protein	#N/A
3061625	RV2748c	DNA translocase ftsK	Cell cycle control, cell division, chromosome partitioning
3073878	RV2764c	thymidylate synthase	Nucleotide transport and metabolism
3077049	RV2768c	PPE family protein	Cell motility
3080805	RV2771c	hypothetical protein	General function prediction only
3086798	IGS	intergenic space or other non-protein-coding region	#N/A
3094861	RV2786c	riboflavin biosynthesis protein RibF	Coenzyme transport and metabolism
3103692	RV2794c	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
3108065	RV2799	hypothetical protein	#N/A
3108684	RV2800	hydrolase	General function prediction only
3111472	RV2802c	hypothetical protein	Function unknown
3111699	RV2802c	hypothetical protein	Function unknown
3118010	RV2812	transposase	Replication, recombination and repair
3119673	IGS	intergenic space or other non-protein-coding region	#N/A
3119701	IGS	intergenic space or other non-protein-coding region	#N/A
3119708	IGS	intergenic space or other non-protein-coding region	#N/A

3119710	IGS	intergenic space or other non-protein-coding region	#N/A
3120189	IGS	intergenic space or other non-protein-coding region	#N/A
3122094	RV2815Ac	hypothetical protein	#N/A
3122106	RV2815Ac	hypothetical protein	#N/A
3133546	RV2825c	hypothetical protein	Function unknown
3136345	RV2828Ac	hypothetical protein	#N/A
3137068	RV2830c	antitoxin	Cell cycle control, cell division, chromosome partitioning
3147180	RV2839c	translation initiation factor IF-2	Translation, ribosomal structure and biogenesis
3162815	RV2853	PE-PGRS family protein	Function unknown
3165084	RV2854	hypothetical protein	Lipid transport and metabolism
3175712	RV2864c	penicillin-binding lipoprotein	Cell wall/membrane/envelope biogenesis
3177894	RV2866	toxin	Translation, ribosomal structure and biogenesis
3178455	IGS	intergenic space or other non-protein-coding region	#N/A
3186870	RV2874	protein dipZ	Post-translational modification, protein turnover, and chaperones
3188596	RV2877c	hypothetical protein	Post-translational modification, protein turnover, and chaperones
3191037	RV2881c	phosphatidate cytidyltransferase	Lipid transport and metabolism
3218353	RV2911	D-alanyl-D-alanine carboxypeptidase	Cell wall/membrane/envelope biogenesis
3226191	RV2916c	signal recognition particle protein	Intracellular trafficking, secretion, and vesicular transport
3228153	RV2917	hypothetical protein	Transcription
3232825	IGS	intergenic space or other non-protein-coding region	#N/A
3236240	RV2922c	chromosome partition protein smc	Cell cycle control, cell division, chromosome partitioning
3243640	IGS	intergenic space or other non-protein-coding region	#N/A
3256504	RV2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism

3260311	RV2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3261390	RV2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3269591	RV2935	non-ribosomal peptide synthetase domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
3270794	RV2935	non-ribosomal peptide synthetase domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
3296853	RV2947c	polyketide synthase 1/15	Secondary metabolites biosynthesis, transport, and catabolism
3302501	IGS	intergenic space or other non-protein-coding region	#N/A
3308616	IGS	intergenic space or other non-protein-coding region	#N/A
3314639	RV2962c	PGL/p-HBAD biosynthesis rhamnosyltransferase	Carbohydrate transport and metabolism
3336209	RV2980	hypothetical protein	#N/A
3336597	IGS	intergenic space or other non-protein-coding region	#N/A
3336835	RV2981c	D-alanine--D-alanine ligase	Cell wall/membrane/envelope biogenesis
3338613	RV2982c	glycerol-3-phosphate dehydrogenase [NAD(P)+]	Energy production and conversion
3351936	RV2994	hypothetical protein	Carbohydrate transport and metabolism
3354906	RV2996c	D-3-phosphoglycerate dehydrogenase	Coenzyme transport and metabolism
3358245	RV2999	lipoprotein LppY	#N/A
3363348	IGS	intergenic space or other non-protein-coding region	#N/A
3367775	RV3009c	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	Translation, ribosomal structure and biogenesis
3402827	RV3042c	phosphoserine phosphatase	Amino acid transport and metabolism
3413590	RV3051c	ribonucleoside-diphosphate reductase subunit alpha	Nucleotide transport and metabolism
3418325	RV3057c	short-chain type alcohol dehydrogenase/reductase	General function prediction only
3418327	RV3057c	short-chain type alcohol dehydrogenase/reductase	General function prediction only
3423181	RV3060c	GntR family transcriptional regulator	Transcription
3425851	RV3062	DNA ligase I	Replication, recombination and repair

3426792	RV3062	DNA ligase I	Replication, recombination and repair
3428914	RV3063	carbon starvation protein	Signal transduction mechanisms
3429199	RV3063	carbon starvation protein	Signal transduction mechanisms
3433229	RV3068c	phosphoglucomutase, alpha-D-glucose phosphate-specific	Carbohydrate transport and metabolism
3440461	RV3077	hydrolase	Inorganic ion transport and metabolism
3440465	RV3077	hydrolase	Inorganic ion transport and metabolism
3454983	RV3088	acyltransferase, WS/DGAT/MGAT	#N/A
3456292	RV3089	o-succinylbenzoate--CoA ligase	Lipid transport and metabolism
3456663	RV3089	o-succinylbenzoate--CoA ligase	Lipid transport and metabolism
3460983	RV3092c	hypothetical protein	Function unknown
3462132	RV3093c	oxidoreductase	Energy production and conversion
3486974	RV3121	cytochrome P450 141 cyp141	Secondary metabolites biosynthesis, transport, and catabolism
3489848	RV3124	hypothetical protein	Signal transduction mechanisms
3491390	RV3125c	PPE family protein	Cell motility
3503892	RV3137	histidinol-phosphatase	Carbohydrate transport and metabolism
3518164	RV3151	NADH-quinone oxidoreductase subunit G	Energy production and conversion
3518552	RV3151	NADH-quinone oxidoreductase subunit G	Energy production and conversion
3537745	RV3169	hypothetical protein	#N/A
3542259	RV3174	oxidoreductase	General function prediction only
3548638	RV3179	ATPase AAA+ superfamily protein	General function prediction only
3556272	RV3190c	hypothetical protein	#N/A
3564397	RV3195	hypothetical protein	Function unknown
3581410	RV3204	methyltransferase	Replication, recombination and repair
3591060	RV3213c	chromosome partitioning protein	Cell cycle control, cell division, chromosome partitioning
3594121	RV3217c	hypothetical protein	#N/A
3604818	RV3228	ribosome small subunit-dependent GTPase A	General function prediction only
3607020	RV3229c	stearoyl-CoA 9-desaturase	Lipid transport and metabolism

3607098	IGS	intergenic space or other non-protein-coding region	#N/A
3610438	RV3234c	acyltransferase, WS/DGAT/MGAT	#N/A
3614979	RV3239c	drug:H <sup>+</sup> antiporter-2 (14 Spanner) (DHA2) family drug resistance MFS transporter	General function prediction only
3622438	RV3243c	hypothetical protein	#N/A
3625062	RV3245c	sensor histidine kinase mtrB	Signal transduction mechanisms
3643389	RV3263	methyltransferase	Replication, recombination and repair
3643627	RV3263	methyltransferase	Replication, recombination and repair
3682155	RV3298c	polyhydroxybutyrate depolymerase	Secondary metabolites biosynthesis, transport, and catabolism
3684086	RV3299c	arylsulfatase	Inorganic ion transport and metabolism
3686203	RV3300c	hypothetical protein	Translation, ribosomal structure and biogenesis
3689520	RV3303c	NAD(P)H dehydrogenase (quinone)	Energy production and conversion
3704767	RV3317	succinate dehydrogenase hydrophobic membrane anchor protein	Energy production and conversion
3710553	RV3324A	4a-hydroxytetrahydrobiopterin dehydratase	#N/A
3714382	RV3328c	sigma-70 family RNA polymerase sigma factor	Transcription
3714928	RV3329	hypothetical protein	#N/A
3717984	RV3331	sugar porter (SP) family MFS transporter	Carbohydrate transport and metabolism
3721977	RV3335c	hypothetical protein	Function unknown
3726058	RV3339Ac	hypothetical protein	#N/A
3731632	RV3343c	PPE family protein	Cell motility
3736799	RV3343c	PPE family protein	Cell motility
3740084	RV3345c	hypothetical protein	#N/A
3740676	RV3345c	hypothetical protein	#N/A
3740684	RV3345c	hypothetical protein	#N/A
3746580	RV3347c	hypothetical protein	Cell motility
3776877	RV3365c	hypothetical protein	Signal transduction mechanisms
3778392	IGS	intergenic space or other non-protein-coding region	#N/A
3795064	IGS	intergenic space or other non-protein-coding region	#N/A
3798275	RV3383c	geranylgeranyl diphosphate synthase, type I	Coenzyme transport and metabolism
3811507	RV3395c	hypothetical protein	#N/A
3813475	RV3396c	GMP synthase	Nucleotide transport and metabolism

3817297	RV3399	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
3823339	RV3403c	hypothetical protein	Function unknown
3825740	RV3406	taurine dioxygenase	Secondary metabolites biosynthesis, transport, and catabolism
3826864	RV3408	toxin	General function prediction only
3829950	RV3410c	IMP dehydrogenase	Nucleotide transport and metabolism
3839051	RV3420c	ribosomal-protein-alanine acetyltransferase	General function prediction only
3839825	RV3421c	hypothetical protein	Post-translational modification, protein turnover, and chaperones
3847864	RV3429/RV3430c	PPE family protein//RNA-directed DNA polymerase	#N/A
3850377	IGS	intergenic space or other non-protein-coding region	#N/A
3859556	RV3439c	hypothetical protein	RNA processing and modification
3860073	RV3440c	hypothetical protein	#N/A
3865174	RV3447c	ESX-4 secretion system protein eccC4	Cell cycle control, cell division, chromosome partitioning
3869250	RV3448	ESX-4 secretion system protein eccD4	#N/A
3873571	RV3451	cutinase	#N/A
3877600	RV3456c	50S ribosomal protein L17	Translation, ribosomal structure and biogenesis
3885085	RV3467	hypothetical protein	#N/A
3886065	RV3468c	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
3892771	RV3476c	MFS transporter, MHS family, alpha-ketoglutarate permease	Carbohydrate transport and metabolism
3892850	RV3476c	MFS transporter, MHS family, alpha-ketoglutarate permease	Carbohydrate transport and metabolism
3894211	IGS	intergenic space or other non-protein-coding region	#N/A
3895870	IGS	intergenic space or other non-protein-coding region	#N/A
3898587	RV3479	hypothetical protein	General function prediction only
3903709	RV3484	hypothetical protein	Transcription
3908266	RV3489	hypothetical protein	#N/A
3928426	RV3507	PE-PGRS family protein	Function unknown
3931075	IGS	intergenic space or other non-protein-coding region	#N/A



3936940	IGS	intergenic space or other non-protein-coding region	#N/A
3941080	RV3511	PE-PGRS family protein	Function unknown
3942660	IGS	intergenic space or other non-protein-coding region	#N/A
3942819	IGS	intergenic space or other non-protein-coding region	#N/A
3943198	IGS	intergenic space or other non-protein-coding region	#N/A
3952979	RV3516	enoyl-CoA hydratase	Lipid transport and metabolism
3958582	RV3521	hypothetical protein	General function prediction only
3959597	RV3522	hypothetical protein	Lipid transport and metabolism
3962617	IGS	intergenic space or other non-protein-coding region	#N/A
3974133	RV3535c	acetaldehyde dehydrogenase	Secondary metabolites biosynthesis, transport, and catabolism
3983450	RV3544c	hypothetical protein	Lipid transport and metabolism
4005786	RV3564	hypothetical protein	Lipid transport and metabolism
4024452	RV3581c	2-C-methyl-D-erythritol 2,4-cyclo diphosphate synthase	Lipid transport and metabolism
4027078	RV3585	DNA repair protein RadA	Post-translational modification, protein turnover, and chaperones
4030722	RV3589	A/G-specific adenine glycosylase	Replication, recombination and repair
4035006	RV3593	hypothetical protein	Defense mechanisms
4037462	RV3595c	PE-PGRS family protein	#N/A
4038466	RV3596c	ATP-dependent Clp protease ATP-binding subunit ClpC	Post-translational modification, protein turnover, and chaperones
4053229	RV3611	hypothetical protein	#N/A
4055980	RV3616c	ESX-1 secretion-associated protein EspA	#N/A
4060083	IGS	intergenic space or other non-protein-coding region	#N/A
4069471	RV3630	hypothetical protein	General function prediction only
4076136	IGS	intergenic space or other non-protein-coding region	#N/A
4088525	RV3648c	CspA family cold shock protein	Transcription
4099459	IGS	intergenic space or other non-protein-coding region	#N/A
4101153	IGS	intergenic space or other non-protein-coding region	#N/A
4111481	RV3669	hypothetical protein	#N/A

4131322	RV3689	hypothetical protein	RNA processing and modification
4135290	RV3693	hypothetical protein	General function prediction only
4136759	RV3694c	hypothetical protein	Function unknown
4150826	RV3707c	hypothetical protein	#N/A
4156277	RV3711c	exonuclease, DNA polymerase III, epsilon subunit	Replication, recombination and repair
4159675	RV3714c	hypothetical protein	#N/A
4162517	RV3719	hypothetical protein	Energy production and conversion
4164122	RV3720	cyclopropane-fatty-acyl-phospholipid synthase	#N/A
4174742	RV3727	oxidoreductase	Function unknown
4182873	RV3731	DNA ligase (ATP)	Replication, recombination and repair
4187663	RV3736	hypothetical protein	Transcription
4187995	RV3737	hypothetical protein	Function unknown
4188834	RV3737	hypothetical protein	Function unknown
4204619	RV3759c	osmoprotectant transporter substrate-binding protein	Cell wall/membrane/envelope biogenesis
4210452	RV3764c	two-component system, OmpR family, sensor kinase	Signal transduction mechanisms
4220352	RV3775	hypothetical protein	Defense mechanisms
4222251	IGS	intergenic space or other non-protein-coding region	#N/A
4223060	RV3777	oxidoreductase	Energy production and conversion
4225144	IGS	intergenic space or other non-protein-coding region	#N/A
4233477	RV3786c	hypothetical protein	Cell wall/membrane/envelope biogenesis
4242360	RV3793	arabinsyltransferase C	#N/A
4242821	RV3793	arabinsyltransferase C	#N/A
4247647	RV3795	arabinsyltransferase B	#N/A
4256100	RV3799c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
4257398	RV3800c	polyketide synthase 13	Secondary metabolites biosynthesis, transport, and catabolism
4264396	RV3802c	hypothetical protein	#N/A
4287189	RV3822	hypothetical protein	#N/A
4291596	RV3823c	membrane transporter mmpL8	General function prediction only
4293250	RV3824c	SL659 acyltransferase papA1	Secondary metabolites biosynthesis, transport, and catabolism

4297763	RV3825c	phthioceranic/hydroxyphthioceranic acid synthase	Secondary metabolites biosynthesis, transport, and catabolism
4299658	RV3825c	phthioceranic/hydroxyphthioceranic acid synthase	Secondary metabolites biosynthesis, transport, and catabolism
4302214	RV3827c	transposase	Replication, recombination and repair
4306333	RV3831	hypothetical protein	#N/A
4307357	RV3833	AraC family transcriptional regulator	Transcription
4315562	RV3842c	glycerophosphoryl diester phosphodiesterase	Energy production and conversion
4316224	RV3843c	hypothetical protein	#N/A
4320228	IGS	intergenic space or other non-protein-coding region	#N/A
4338909	IGS	intergenic space or other non-protein-coding region	#N/A
4340507	RV3864	hypothetical protein	#N/A
4341149	RV3864	hypothetical protein	#N/A
4351216	RV3872	PE family protein	#N/A
4356287	RV3877	ESX-1 secretion system protein eccD1	#N/A
4363246	RV3882c	ESX-1 secretion system protein EccE1	#N/A
4366449	RV3884c	ESX-2 secretion system protein EccA2	Post-translational modification, protein turnover, and chaperones
4373673	RV3889c	ESX-2 secretion-associated protein EspG2	#N/A
4375805	RV3892c	PPE family protein	#N/A
4376275	IGS	intergenic space or other non-protein-coding region	#N/A
4379544	RV3894c	type VII secretion protein EccCa	Cell cycle control, cell division, chromosome partitioning
4379857	RV3894c	type VII secretion protein EccCa	Cell cycle control, cell division, chromosome partitioning
4382151	RV3896c	hypothetical protein	General function prediction only
4382231	RV3896c	hypothetical protein	General function prediction only
4382452	RV3896c	hypothetical protein	General function prediction only
4385707	RV3900c	hypothetical protein	#N/A
4386789	RV3901c	hypothetical protein	#N/A
4390930	RV3905c	WXG100 family type VII secretion target	Function unknown
4391730	RV3906c	hypothetical protein	#N/A
4400035	RV3910	integral membrane protein MviN	General function prediction only
4405133	RV3916c	hypothetical protein	#N/A

4408128	RV3919c	ribosomal RNA small subunit methyltransferase G	#N/A
4408332	RV3919c	ribosomal RNA small subunit methyltransferase G	#N/A

Tabla Anexo 5. Variantes genetica de *Mycobacterium tuberculosis* INS\_XDR

Posición	Gene_ID	Producto	COG
1977	IGS	intergenic space or other non-protein-coding region	#N/A
4013	RV0003	DNA replication and repair protein RecF	Replication, recombination and repair
7362	RV0006	DNA gyrase subunit A	Replication, recombination and repair
7582	RV0006	DNA gyrase subunit A	Replication, recombination and repair
7585	RV0006	DNA gyrase subunit A	Replication, recombination and repair
8040	RV0006	DNA gyrase subunit A	Replication, recombination and repair
9304	RV0006	DNA gyrase subunit A	Replication, recombination and repair
11879	RV0008c	membrane protein	#N/A
14251	RV0012	hypothetical protein	Function unknown
17608	RV0015c	serine/threonine protein kinase	General function prediction only
21795	RV0018c	PP2C-family Ser/Thr phosphatase	Signal transduction mechanisms
23681	RV0019c	hypothetical protein	Signal transduction mechanisms
26959	IGS	intergenic space or other non-protein-coding region	#N/A
27463	IGS	intergenic space or other non-protein-coding region	#N/A
33457	RV0030	hypothetical protein	#N/A
33551	RV0030	hypothetical protein	#N/A
34044	IGS	intergenic space or other non-protein-coding region	#N/A
37031	RV0034	hypothetical protein	General function prediction only
37763	RV0035	fatty-acid-CoA ligase FadD34	Lipid transport and metabolism
42967	RV0040c	secreted proline rich protein mtc28	#N/A
55208	RV0050	peptidoglycan glycosyltransferase	Cell wall/membrane/envelope biogenesis
55250	RV0050	peptidoglycan glycosyltransferase	Cell wall/membrane/envelope biogenesis
60063	RV0057	hypothetical protein	#N/A
62046	RV0058	replicative DNA helicase	Replication, recombination and repair
63768	RV0059	hypothetical protein	#N/A
66357	RV0062	endoglucanase	Carbohydrate transport and metabolism
70813	RV0064	hypothetical protein	Function unknown
71333	RV0064	hypothetical protein	Function unknown
74056	RV0066c	isocitrate dehydrogenase	Energy production and conversion
75937	RV0068	oxidoreductase	Lipid transport and metabolism
80613	IGS	intergenic space or other non-protein-coding region	#N/A
92196	RV0083	oxidoreductase	Energy production and conversion
93141	RV0084	formate hydrogenlyase hycD	Energy production and conversion
105042	RV0095c	hypothetical protein	#N/A
122106	RV0103c	cation-transporting P-type ATPase B	Inorganic ion transport and metabolism
128001	RV0107c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
130224	RV0107c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
133837	IGS	intergenic space or other non-protein-coding region	#N/A
133860	IGS	intergenic space or other non-protein-coding region	#N/A
146085	RV0120c	elongation factor G-like protein	Translation, ribosomal structure and biogenesis
149321	IGS	intergenic space or other non-protein-coding region	#N/A

154189	IGS	intergenic space or other non-protein-coding region	#N/A
154281	RV0127	hypothetical protein	Carbohydrate transport and metabolism
157290	RV0129c	antigen 85-C	General function prediction only
161085	RV0133	acetyltransferase	General function prediction only
162579	RV0134	epoxide hydrolase EphF	General function prediction only
177855	RV0151c	PE family protein	#N/A
178939	RV0151c	PE family protein	#N/A
180023	RV0152c	PE family protein	#N/A
182220	RV0154c	acyl-CoA dehydrogenase	Lipid transport and metabolism
184459	RV0155	NAD(P)(+) transhydrogenase (AB-specific), alpha subunit	Energy production and conversion
184762	RV0156	NAD(P) transhydrogenase subunit alpha	Energy production and conversion
188798	RV0159c	PE family protein	#N/A
194679	RV0165c	GntR family transcriptional regulator	#N/A
196520	RV0166	fatty-acyl-CoA synthase	Lipid transport and metabolism
196640	RV0166	fatty-acyl-CoA synthase	Lipid transport and metabolism
200152	RV0170	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
203267	RV0172	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
206337	RV0174	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
207224	RV0175	MCE-associated membrane protein	#N/A
212351	RV0181c	hypothetical protein	General function prediction only
223940	RV0192	hypothetical protein	Function unknown
225321	RV0193c	hypothetical protein	#N/A
227096	RV0194	ATP-binding cassette, subfamily C	Defense mechanisms
231112	RV0195	two component system transcriptional regulator, luxR-family	Signal transduction mechanisms
234448	RV0197	nitrate reductase catalytic subunit	Energy production and conversion
261869	RV0218	hypothetical protein	General function prediction only
265554	RV0222	enoyl-CoA hydratase	Lipid transport and metabolism
278681	RV0233	R2-like ligand binding oxidase	Nucleotide transport and metabolism
283614	RV0236c	hypothetical protein	#N/A
285772	RV0236c	hypothetical protein	#N/A
285871	RV0236c	hypothetical protein	#N/A
294586	RV0244c	acyl-CoA dehydrogenase fadE5	Lipid transport and metabolism
305362	RV0252	nitrite reductase [NAD(P)H], large subunit	Energy production and conversion
310973	RV0259c	hypothetical protein	Function unknown
311613	RV0260c	uroporphyrinogen-III synthase	Coenzyme transport and metabolism
316129	RV0264c	hypothetical protein	Amino acid transport and metabolism
327897	RV0271c	acyl-CoA dehydrogenase fadE6	Lipid transport and metabolism
333406	IGS	intergenic space or other non-protein-coding region	#N/A
333423	IGS	intergenic space or other non-protein-coding region	#N/A
333430	IGS	intergenic space or other non-protein-coding region	#N/A
333437	RV0278c	PE-PGRS family protein	#N/A
333440	RV0278c	PE-PGRS family protein	#N/A
336682	RV0279c	WAG22 antigen	#N/A

336688	RV0279c	WAG22 antigen	#N/A
336692	RV0279c	WAG22 antigen	#N/A
337820	RV0279c	WAG22 antigen	#N/A
340372	RV0280	PPE family protein	Cell motility
346275	RV0284	ESX-3 secretion system protein eccC3	Cell cycle control, cell division, chromosome partitioning
356528	RV0292	ESX-3 secretion system protein EccE3	#N/A
376774	RV0307c	hypothetical protein	#N/A
384380	RV0315	hypothetical protein	Carbohydrate transport and metabolism
386432	RV0318c	ZIP family zinc transporter	Inorganic ion transport and metabolism
387353	RV0319	pyrrolidone-carboxylate peptidase	Post-translational modification, protein turnover, and chaperones
398116	RV0332	hypothetical protein	#N/A
403364	RV0338c	oxidoreductase	Energy production and conversion
403920	RV0338c	oxidoreductase	Energy production and conversion
403980	RV0338c	oxidoreductase	Energy production and conversion
404326	RV0338c	oxidoreductase	Energy production and conversion
408723	RV0340	hypothetical protein	#N/A
414486	RV0344c	lipoprotein LpqJ	Lipid transport and metabolism
420008	RV0350	chaperone dnaK	Post-translational modification, protein turnover, and chaperones
425343	RV0355c	PPE family protein	#N/A
440561	RV0362	magnesium transporter	Inorganic ion transport and metabolism
444352	RV0366c	hypothetical protein	Function unknown
457453	RV0381c	hypothetical protein	#N/A
463337	IGS	intergenic space or other non-protein-coding region	#N/A
472708	IGS	intergenic space or other non-protein-coding region	#N/A
475181	RV0395	hypothetical protein	#N/A
479635	RV0400c	glutaryl-CoA dehydrogenase	Lipid transport and metabolism
489938	RV0405/RV0406c	polyketide synthase 12/beta lactamase like protein	#N/A
498560	RV0412c	hypothetical protein	#N/A
501538	RV0415	glycine oxidase ThiO	Amino acid transport and metabolism
502592	RV0417	thiazole synthase	Coenzyme transport and metabolism
503357	IGS	intergenic space or other non-protein-coding region	#N/A
513260	RV0425c	HAD ATPase, P-type, family IC	Inorganic ion transport and metabolism
534694	RV0446c	hypothetical protein	Function unknown
535698	RV0447c	cyclopropane-fatty-acyl-phospholipid synthase	Cell wall/membrane/envelope biogenesis
541204	RV0450c	transmembrane transporter mmpL4	General function prediction only
551528	RV0459	hypothetical protein	Function unknown
558504	RV0467	isocitrate lyase	Energy production and conversion
563423	RV0472c	TetR family transcriptional regulator	Transcription
573265	RV0484c	oxidoreductase	General function prediction only
584441	RV0493c	hypothetical protein	#N/A
586217	RV0495c	hypothetical protein	#N/A
589539	RV0499	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
590439	RV0500	pyrroline-5-carboxylate reductase	Amino acid transport and metabolism

597819	RV0507	transmembrane transporter mmpL2	General function prediction only
598478	RV0507	transmembrane transporter mmpL2	General function prediction only
610123	IGS	intergenic space or other non-protein-coding region	#N/A
630725	RV0538	hypothetical protein	General function prediction only
635636	RV0543c	hypothetical protein	#N/A
637322	RV0545c	PIT family inorganic phosphate transporter	Inorganic ion transport and metabolism
637925	RV0546c	lactoylglutathione lyase	General function prediction only
648005	RV0556	hypothetical protein	#N/A
664252	RV0571c	hypothetical protein	General function prediction only
665296	RV0572c	hypothetical protein	#N/A
669401	RV0575c	oxidoreductase	Coenzyme transport and metabolism
673241	RV0578c	PE-PGRS family protein	#N/A
675676	RV0578c	PE-PGRS family protein	#N/A
675800	RV0578c	PE-PGRS family protein	#N/A
685464	RV0587	ABC transporter permease	Secondary metabolites biosynthesis, transport, and catabolism
685611	RV0587	ABC transporter permease	Secondary metabolites biosynthesis, transport, and catabolism
686975	RV0589	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
690468	RV0591	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
697117	RV0597A	hypothetical protein	#N/A
698971	RV0601c	sensor histidine kinase component HK2	Signal transduction mechanisms
709229	RV0613c	hypothetical protein	#N/A
726706	RV0631c	exodeoxyribonuclease V subunit gamma	Replication, recombination and repair
736713	RV0642c	hydroxymycolate synthase MmaA4	Cell wall/membrane/envelope biogenesis
752346	RV0655	ABC transporter ATP-binding protein	Secondary metabolites biosynthesis, transport, and catabolism
752805	IGS	intergenic space or other non-protein-coding region	#N/A
754189	RV0658c	hypothetical protein	General function prediction only
761113	RV0667	DNA-directed RNA polymerase subunit beta	Transcription
764998	RV0668	DNA-directed RNA polymerase subunit beta'	Transcription
769666	IGS	intergenic space or other non-protein-coding region	#N/A
773812	RV0673	enoyl-CoA hydratase	Lipid transport and metabolism
775642	RV0676c	transmembrane transporter mmpL5	General function prediction only
781398	IGS	intergenic space or other non-protein-coding region	#N/A
796512	RV0696	mycofactocin system glycosyltransferase	Cell wall/membrane/envelope biogenesis
812811	RV0717	30S ribosomal protein S14	Translation, ribosomal structure and biogenesis
820486	RV0727c	L-ribulose-5-phosphate 4-epimerase	Carbohydrate transport and metabolism
826094	RV0732	preprotein translocase subunit secY	Intracellular trafficking, secretion, and vesicular transport
839952	RV0747	PE-PGRS family protein	#N/A
841767	RV0749A	hypothetical protein	#N/A
852913	RV0758	two-component system, OmpR family, sensor kinase	Signal transduction mechanisms
857699	RV0764c	lanosterol 14-alpha demethylase	Secondary metabolites biosynthesis, transport, and catabolism
862990	RV0769	3-oxoacyl-ACP reductase	Lipid transport and metabolism
863808	RV0770	dehydrogenase/reductase	Lipid transport and metabolism



880565	RV0785	KsdD-like steroid dehydrogenase	General function prediction only
882260	RV0787	hypothetical protein	#N/A
885545	RV0791c	hypothetical protein	Energy production and conversion
893736	RV0800	aminopeptidase	Amino acid transport and metabolism
900224	RV0806c	exopolysaccharide phosphotransferase CpsY	#N/A
903553	RV0808	amidophosphoribosyltransferase	Nucleotide transport and metabolism
903916	RV0809	phosphoribosylformylglycinamide cyclo-ligase	Nucleotide transport and metabolism
906860	RV0812	4-amino-4-deoxychorismate lyase	Amino acid transport and metabolism
907819	RV0813c	hypothetical protein	#N/A
919396	RV0825c	hypothetical protein	Transcription
927388	RV0833	PE-PGRS family protein	Function unknown
942482	IGS	intergenic space or other non-protein-coding region	#N/A
945217	RV0848	cysteine synthase A	Amino acid transport and metabolism
947432	RV0850	transposase	#N/A
949538	RV0853c	alpha-keto-acid decarboxylase	Carbohydrate transport and metabolism
955014	IGS	intergenic space or other non-protein-coding region	#N/A
955527	RV0859	acetyl-CoA C-acetyltransferase	Lipid transport and metabolism
958925	RV0861c	DNA excision repair protein ERCC-3	#N/A
979707	RV0881	methyltransferase	Translation, ribosomal structure and biogenesis
986466	IGS	intergenic space or other non-protein-coding region	#N/A
989722	RV0889c	citrate synthase	Energy production and conversion
991518	RV0890c	LuxR family transcriptional regulator	General function prediction only
993349	RV0891c	transcriptional regulator	Signal transduction mechanisms
996200	RV0893c	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
997902	RV0895	acyltransferase, WS/DGAT/MGAT	#N/A
1029253	RV0922	transposase	Replication, recombination and repair
1037016	RV0930	phosphate transporter permease pstA 1	Inorganic ion transport and metabolism
1040054	RV0932c	phosphate-binding protein pstS 2	Inorganic ion transport and metabolism
1042229	RV0934	phosphate-binding protein pstS 1	Inorganic ion transport and metabolism
1047169	RV0938	DNA ligase D	Replication, recombination and repair
1054366	RV0945	oxidoreductase	General function prediction only
1057792	RV0948c	chorismate mutase	Amino acid transport and metabolism
1064520	RV0953c	oxidoreductase	Energy production and conversion
1068155	RV0956	phosphoribosylglycinamide formyltransferase	Nucleotide transport and metabolism
1068436	RV0957	bifunctional purine biosynthesis protein purH	Nucleotide transport and metabolism
1069454	RV0957	bifunctional purine biosynthesis protein purH	Nucleotide transport and metabolism
1070706	RV0958	magnesium chelatase subunit I	Coenzyme transport and metabolism
1073037	RV0959	hypothetical protein	General function prediction only
1074562	RV0962c	lipoprotein LprP	#N/A
1075283	IGS	intergenic space or other non-protein-coding region	#N/A
1076313	RV0964c	hypothetical protein	#N/A
1077316	RV0966c	hypothetical protein	#N/A
1079931	RV0969	heavy metal translocating P-type ATPase	#N/A
1081685	RV0970	hypothetical protein	#N/A

1087197	RV0974c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
1087283	RV0974c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
1091815	RV0977	PE-PGRS family protein	Function unknown
1093326	IGS	intergenic space or other non-protein-coding region	#N/A
1096636	IGS	intergenic space or other non-protein-coding region	#N/A
1100238	RV0983	serine protease PepD	Post-translational modification, protein turnover, and chaperones
1103253	RV0987	ABC transporter permease	Cell wall/membrane/envelope biogenesis
1106426	RV0989c	heptaprenyl diphosphate synthase	Coenzyme transport and metabolism
1107438	IGS	intergenic space or other non-protein-coding region	#N/A
1107921	RV0990c	hypothetical protein	Cell motility
1109979	RV0993	UTP-glucose-1-phosphate uridylyltransferase	Cell wall/membrane/envelope biogenesis
1126893	RV1007c	methionyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1127652	RV1008	TatD family hydrolase	Replication, recombination and repair
1142270	RV1020	transcription-repair-coupling factor	Replication, recombination and repair
1144668	RV1023	enolase	Carbohydrate transport and metabolism
1149555	RV1028c	sensor protein kdpD	Signal transduction mechanisms
1150589	RV1028c	sensor protein kdpD	Signal transduction mechanisms
1162278	RV1039c	PPE family protein	Cell motility
1163138	RV1040c	PE family protein	#N/A
1164575	IGS	intergenic space or other non-protein-coding region	#N/A
1167651	RV1044	hypothetical protein	Transcription
1177999	RV1056	hypothetical protein	Function unknown
1178121	RV1056	hypothetical protein	Function unknown
1190159	RV1067c	PE-PGRS family protein	Function unknown
1199552	RV1075c	hypothetical protein	Amino acid transport and metabolism
1200423	IGS	intergenic space or other non-protein-coding region	#N/A
1212081	RV1087	PE-PGRS family protein	Function unknown
1212331	RV1087	PE-PGRS family protein	Function unknown
1212340	RV1087	PE-PGRS family protein	Function unknown
1218306	RV1091	PE-PGRS family protein	#N/A
1218742	RV1091	PE-PGRS family protein	#N/A
1220685	RV1093	serine hydroxymethyltransferase 1	#N/A
1224372	IGS	intergenic space or other non-protein-coding region	#N/A
1248983	RV1125	hypothetical protein	#N/A
1272015	IGS	intergenic space or other non-protein-coding region	#N/A
1281123	RV1154c	hypothetical protein	Function unknown
1292107	RV1162	nitrate reductase subunit beta	Energy production and conversion
1307603	RV1175c	2,4-dienoyl-CoA reductase	Energy production and conversion
1324730	RV1184c	hypothetical protein	#N/A
1327896	RV1186c	hypothetical protein	Signal transduction mechanisms
1328693	RV1186c	hypothetical protein	Signal transduction mechanisms
1349007	RV1205	hypothetical protein	General function prediction only
1354443	RV1211	hypothetical protein	#N/A
1360215	RV1217c	antibiotic transporter permease	Cell wall/membrane/envelope biogenesis

1363365	RV1219c	transcriptional regulator	Transcription
1372893	RV1229c	ATP-binding protein	Cell cycle control, cell division, chromosome partitioning
1374071	RV1230c	membrane protein	Cell wall/membrane/envelope biogenesis
1375730	RV1232c	hypothetical protein	Inorganic ion transport and metabolism
1382634	RV1239c	magnesium and cobalt transporter CorA	Inorganic ion transport and metabolism
1389744	RV1248c	2-oxoglutarate decarboxylase	Energy production and conversion
1393632	RV1249c	membrane protein	#N/A
1396928	RV1251c	hypothetical protein	General function prediction only
1404175	RV1256c	cytochrome P450 130 cyp130	Secondary metabolites biosynthesis, transport, and catabolism
1411216	RV1263	amidase	Translation, ribosomal structure and biogenesis
1413154	IGS	intergenic space or other non-protein-coding region	#N/A
1428512	RV1278	hypothetical protein	Function unknown
1431811	RV1280c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
1440475	RV1286	bifunctional enzyme CysN/CysC	Inorganic ion transport and metabolism
1445480	IGS	intergenic space or other non-protein-coding region	#N/A
1445787	RV1291c	hypothetical protein	#N/A
1446929	RV1292	arginyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
1452077	RV1296	homoserine kinase	Amino acid transport and metabolism
1457150	RV1300	HemK family methyltransferase	Translation, ribosomal structure and biogenesis
1480030	RV1318c	adenylate cyclase	Signal transduction mechanisms
1480978	RV1319c	adenylate cyclase	Signal transduction mechanisms
1482633	RV1320c	adenylate cyclase	Signal transduction mechanisms
1484714	RV1321	hypothetical protein	Replication, recombination and repair
1499280	RV1330c	nicotinate phosphoribosyltransferase pncB1	Coenzyme transport and metabolism
1512037	RV1347c	lysine N-acyltransferase mbtK	Translation, ribosomal structure and biogenesis
1519853	RV1353c	transcriptional regulator	Transcription
1523309	RV1355c	molybdopterin biosynthesis protein MoeY	Coenzyme transport and metabolism
1526825	RV1358	transcriptional regulator	General function prediction only
1532302	RV1360	oxidoreductase	Energy production and conversion
1535147	RV1363c	membrane protein	#N/A
1536257	RV1364c	anti-anti-sigma factor	Signal transduction mechanisms
1546536	RV1373	glycolipid sulfotransferase	#N/A
1547131	RV1374c	hypothetical protein	#N/A
1552553	RV1378c	hypothetical protein	#N/A
1563723	RV1388	integration host factor mihF	#N/A
1570572	RV1394c	cytochrome P450 132 cyp132	Secondary metabolites biosynthesis, transport, and catabolism
1573021	RV1396c	PE-PGRS family protein	Cell wall/membrane/envelope biogenesis
1573666	RV1396c	PE-PGRS family protein	Cell wall/membrane/envelope biogenesis
1575201	RV1399c	lipase lipH	Lipid transport and metabolism
1586255	RV1410c	MFS-type drug efflux transporter P55	Carbohydrate transport and metabolism
1588905	RV1412	riboflavin synthase subunit alpha	Coenzyme transport and metabolism
1609846	RV1431	hypothetical protein	#N/A

1612295	RV1434	hypothetical protein	#N/A
1613041	IGS	intergenic space or other non-protein-coding region	#N/A
1613966	RV1436	glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate transport and metabolism
1618635	RV1441c	triacylglycerol lipase	#N/A
1618640	RV1441c	triacylglycerol lipase	#N/A
1618653	RV1441c	triacylglycerol lipase	#N/A
1618667	RV1441c	triacylglycerol lipase	#N/A
1618679	RV1441c	triacylglycerol lipase	#N/A
1618687	RV1441c	triacylglycerol lipase	#N/A
1624797	RV1446c	opcA protein	Carbohydrate transport and metabolism
1628595	RV1449c	transketolase	Carbohydrate transport and metabolism
1630154	RV1449c	transketolase	Carbohydrate transport and metabolism
1636832	RV1452c	PE-PGRS family protein	#N/A
1636924	RV1452c	PE-PGRS family protein	#N/A
1636933	RV1452c	PE-PGRS family protein	#N/A
1636940	RV1452c	PE-PGRS family protein	#N/A
1636942	RV1452c	PE-PGRS family protein	#N/A
1636986	RV1452c	PE-PGRS family protein	#N/A
1639600	RV1453	hypothetical protein	Signal transduction mechanisms
1645808	RV1459c	hypothetical protein	#N/A
1650078	RV1462	Fe-S cluster assembly protein SufD	Post-translational modification, protein turnover, and chaperones
1651312	RV1463	FeS assembly ATPase SufC	Post-translational modification, protein turnover, and chaperones
1654054	RV1467c	acyl-CoA dehydrogenase fadE15	Lipid transport and metabolism
1676296	RV1486c	hypothetical protein	#N/A
1689355	RV1498c	methyltransferase	Coenzyme transport and metabolism
1692147	RV1501	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
1692801	IGS	intergenic space or other non-protein-coding region	#N/A
1693567	RV1502	hypothetical protein	#N/A
1698917	RV1508c	membrane protein	#N/A
1706125	RV1514c	hypothetical protein	Cell wall/membrane/envelope biogenesis
1711633	RV1520	sugar transferase	#N/A
1717687	RV1523	methyltransferase	Coenzyme transport and metabolism
1718767	RV1524	glycosyltransferase	Carbohydrate transport and metabolism
1728843	IGS	intergenic space or other non-protein-coding region	#N/A
1732245	RV1530	alcohol dehydrogenase	Amino acid transport and metabolism
1735000	RV1534	transcriptional regulator	Transcription
1740777	RV1537	DNA polymerase IV 1	#N/A
1752567	RV1548c	PPE family protein	Cell motility
1755525	RV1551	glycerol-3-phosphate O-acyltransferase	Lipid transport and metabolism
1759258	RV1552	fumarate reductase flavoprotein subunit	Energy production and conversion
1760298	RV1554	fumarate reductase subunit C	Energy production and conversion
1762621	RV1557	hypothetical protein	General function prediction only
1773693	RV1565c	hypothetical protein	Lipid transport and metabolism

1775318	IGS	intergenic space or other non-protein-coding region	#N/A
1777188	RV1569	8-amino-7-oxononanoate synthase	Coenzyme transport and metabolism
1777219	RV1569	8-amino-7-oxononanoate synthase	Coenzyme transport and metabolism
1778436	RV1570	dethiobiotin synthetase	Coenzyme transport and metabolism
1779249	RV1572c	hypothetical protein	#N/A
1782625	IGS	intergenic space or other non-protein-coding region	#N/A
1782717	IGS	intergenic space or other non-protein-coding region	#N/A
1782774	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782780	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782791	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782793	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1782799	RV1578c	phiRv1 phage protein	Replication, recombination and repair
1793776	IGS	intergenic space or other non-protein-coding region	#N/A
1797034	RV1595	L-aspartate oxidase	Coenzyme transport and metabolism
1798362	RV1597	hypothetical protein	Coenzyme transport and metabolism
1803272	RV1602	imidazole glycerol phosphate synthase subunit HisH	Amino acid transport and metabolism
1804416	RV1604	myo-inositol-1(or 4)-monophosphatase	Carbohydrate transport and metabolism
1805955	RV1606	phosphoribosyl-AMP cyclohydrolase	Amino acid transport and metabolism
1815611	RV1615	hypothetical protein	Function unknown
1817983	RV1618	acyl-CoA thioesterase II	Lipid transport and metabolism
1819598	RV1619	hypothetical protein	Function unknown
1836293	IGS	intergenic space or other non-protein-coding region	#N/A
1847926	RV1639c	hypothetical protein	General function prediction only
1854307	RV1644	TrmH family RNA methyltransferase	Translation, ribosomal structure and biogenesis
1856784	RV1647	adenylate cyclase	Signal transduction mechanisms
1885779	RV1662	polyketide synthase 8	Secondary metabolites biosynthesis, transport, and catabolism
1895181	RV1668c	hypothetical protein	General function prediction only
1895524	IGS	intergenic space or other non-protein-coding region	#N/A
1899358	RV1674c	transcriptional regulator	Inorganic ion transport and metabolism
1901500	RV1676	hypothetical protein	Post-translational modification, protein turnover, and chaperones
1907303	RV1682	hypothetical protein	#N/A
1917979	RV1694	hemolysin TlyA family protein	Translation, ribosomal structure and biogenesis
1931186	RV1704c	AAT family amino acid transporter	Amino acid transport and metabolism
1933995	RV1706Xc	hypothetical protein	#N/A
1938034	RV1709	segregation and condensation protein A	Function unknown
1943046	RV1715	3-hydroxybutyryl-CoA dehydrogenase	Lipid transport and metabolism
1943599	RV1716	arylformamidase	General function prediction only
1944114	RV1716	arylformamidase	General function prediction only
1944409	RV1716	arylformamidase	General function prediction only
1945135	RV1718	hypothetical protein	Function unknown
1947910	RV1722	carboxylase	Lipid transport and metabolism
1950774	RV1724c	hypothetical protein	#N/A
1960291	RV1733c	hypothetical protein	#N/A

1967244	RV1739c	sulfate permease	Inorganic ion transport and metabolism
1981063	IGS	intergenic space or other non-protein-coding region	#N/A
1982908	RV1753c	PPE family protein	Cell motility
1983320	RV1753c	PPE family protein	Cell motility
1989086	RV1758	esterase/lipase	#N/A
1990926	RV1759c	WAG22 antigen	#N/A
1993815	RV1760	acyltransferase, WS/DGAT/MGAT	#N/A
1993937	RV1760	acyltransferase, WS/DGAT/MGAT	#N/A
2006039	RV1771	FAD-linked oxidoreductase	Energy production and conversion
2007552	RV1773c	hypothetical protein	Transcription
2022875	RV1784	ESX-5 secretion system protein eccCa5	#N/A
2024061	RV1785c	cytochrome P450 143 cyp143	Secondary metabolites biosynthesis, transport, and catabolism
2037723	RV1798	ESX-5 secretion system protein EccA5	General function prediction only
2044198	RV1802	PPE family protein	Cell motility
2045317	RV1803c	hypothetical protein	#N/A
2048986	RV1807	hypothetical protein	Cell motility
2049072	RV1807	hypothetical protein	Cell motility
2049104	RV1807	hypothetical protein	Cell motility
2052042	RV1809	hypothetical protein	Cell motility
2053418	IGS	intergenic space or other non-protein-coding region	#N/A
2053689	RV1811	Mg2+ transporter-C (MgtC) family protein	Function unknown
2055278	RV1812c	NADH dehydrogenase-like protein	Energy production and conversion
2071583	RV1825	hypothetical protein	Function unknown
2074516	IGS	intergenic space or other non-protein-coding region	#N/A
2074572	IGS	intergenic space or other non-protein-coding region	#N/A
2077260	RV1832	glycine dehydrogenase	Amino acid transport and metabolism
2096193	RV1846c	transcriptional regulator Blal	Transcription
2097997	RV1850	urease subunit alpha	Amino acid transport and metabolism
2108148	RV1860	alanine and proline-rich secreted protein apa	Function unknown
2108845	IGS	intergenic space or other non-protein-coding region	#N/A
2116910	RV1867	acetyl-CoA C-acetyltransferase	Lipid transport and metabolism
2123176	IGS	intergenic space or other non-protein-coding region	#N/A
2123189	RV1873	hypothetical protein	Function unknown
2128877	RV1878	glutamine synthetase	Amino acid transport and metabolism
2128915	RV1878	glutamine synthetase	Amino acid transport and metabolism
2134222	RV1884c	hypothetical protein	#N/A
2135877	IGS	intergenic space or other non-protein-coding region	#N/A
2137549	RV1888c	transmembrane protein	#N/A
2143335	RV1895	chlorophyll synthesis pathway, bchC	Amino acid transport and metabolism
2144624	RV1897c	D-tyrosyl-tRNA(Tyr) deacylase	Translation, ribosomal structure and biogenesis
2147029	RV1900c	hypothetical protein	Signal transduction mechanisms
2153417	RV1907c	hypothetical protein	#N/A
2155175	RV1908c	catalase-peroxidase	Inorganic ion transport and metabolism
2162188	RV1916	isocitrate lyase	Energy production and conversion

2163797	RV1917c	hypothetical protein	Cell motility
2165293	RV1917c	hypothetical protein	Cell motility
2165561	RV1917c	hypothetical protein	Cell motility
2169005	RV1918c	PPE family protein	Cell motility
2171973	RV1920	membrane protein	Lipid transport and metabolism
2174223	RV1922	hypothetical protein	Defense mechanisms
2180381	RV1927	hypothetical protein	Function unknown
2207532	IGS	intergenic space or other non-protein-coding region	#N/A
2211834	RV1968	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2212881	RV1969	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2216256	RV1971	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2216451	RV1971	virulence factor Mce family protein	Secondary metabolites biosynthesis, transport, and catabolism
2220520	RV1977	hypothetical protein	Post-translational modification, protein turnover, and chaperones
2223301	IGS	intergenic space or other non-protein-coding region	#N/A
2228975	IGS	intergenic space or other non-protein-coding region	#N/A
2231140	RV1987	chitinase	Carbohydrate transport and metabolism
2245540	RV2000	hypothetical protein	Energy production and conversion
2260533	IGS	intergenic space or other non-protein-coding region	#N/A
2264790	RV2017	transcriptional regulator	Amino acid transport and metabolism
2265067	IGS	intergenic space or other non-protein-coding region	#N/A
2267380	RV2022c	hypothetical protein	Function unknown
2269788	RV2024c	hypothetical protein	General function prediction only
2270110	RV2024c	hypothetical protein	General function prediction only
2278454	RV2030c	erythromycin esterase	General function prediction only
2283038	RV2037c	hypothetical protein	General function prediction only
2285259	RV2039c	multiple sugar transporter permease	Carbohydrate transport and metabolism
2286747	RV2041c	multiple sugar transporter substrate-binding protein	Carbohydrate transport and metabolism
2287129	RV2041c	multiple sugar transporter substrate-binding protein	Carbohydrate transport and metabolism
2288470	RV2042c	hypothetical protein	#N/A
2289221	RV2043c	nicotinamidase/pyrazinamidase	Secondary metabolites biosynthesis, transport, and catabolism
2300245	RV2048c	polyketide synthase 12	#N/A
2300560	RV2048c	polyketide synthase 12	#N/A
2300563	RV2048c	polyketide synthase 12	#N/A
2306314	RV2048c	polyketide synthase 12	#N/A
2306526	RV2048c	polyketide synthase 12	#N/A
2306626	RV2048c	polyketide synthase 12	#N/A
2306629	RV2048c	polyketide synthase 12	#N/A
2306638	RV2048c	polyketide synthase 12	#N/A
2306640	RV2048c	polyketide synthase 12	#N/A
2306650	RV2048c	polyketide synthase 12	#N/A
2306668	RV2048c	polyketide synthase 12	#N/A
2306674	RV2048c	polyketide synthase 12	#N/A

2306680	RV2048c	polyketide synthase 12	#N/A
2306689	RV2048c	polyketide synthase 12	#N/A
2306692	RV2048c	polyketide synthase 12	#N/A
2306698	RV2048c	polyketide synthase 12	#N/A
2306704	RV2048c	polyketide synthase 12	#N/A
2306707	RV2048c	polyketide synthase 12	#N/A
2306713	RV2048c	polyketide synthase 12	#N/A
2306715	RV2048c	polyketide synthase 12	#N/A
2306724	RV2048c	polyketide synthase 12	#N/A
2310578	RV2051c	apolipoprotein N-acyltransferase	Cell wall/membrane/envelope biogenesis
2315677	RV2059	zinc/manganese transporter substrate-binding protein	Inorganic ion transport and metabolism
2329541	RV2072c	precorrin-6Y C(5,15)-methyltransferase	Coenzyme transport and metabolism
2334015	RV2077c	hypothetical protein	#N/A
2335083	RV2078	hypothetical protein	#N/A
2335502	RV2079	hypothetical protein	Cell wall/membrane/envelope biogenesis
2340629	RV2082	hypothetical protein	#N/A
2341644	RV2083	hypothetical protein	#N/A
2345045	RV2088	serine/threonine protein kinase	General function prediction only
2347450	IGS	intergenic space or other non-protein-coding region	#N/A
2348454	RV2090	5'-3' exonuclease	Replication, recombination and repair
2353421	RV2095c	protein pafC	Transcription
2361612	RV2101	helicase helZ	#N/A
2362049	RV2101	helicase helZ	#N/A
2386397	RV2125	hypothetical protein	General function prediction only
2401878	IGS	intergenic space or other non-protein-coding region	#N/A
2401891	IGS	intergenic space or other non-protein-coding region	#N/A
2415664	RV2155c	UDP-N-acetylmuramoylalanine--D-glutamate ligase	#N/A
2424933	IGS	intergenic space or other non-protein-coding region	#N/A
2437979	RV2176	serine/threonine protein kinase	General function prediction only
2440934	RV2178c	3-deoxy-7-phosphoheptulonate synthase	Amino acid transport and metabolism
2451954	RV2189c	hypothetical protein	#N/A
2462879	RV2198c	membrane protein mmpS3	General function prediction only
2463102	IGS	intergenic space or other non-protein-coding region	#N/A
2478647	RV2213	leucyl aminopeptidase	Amino acid transport and metabolism
2480166	RV2214c	oxidoreductase	General function prediction only
2484263	RV2216	epimerase	General function prediction only
2499734	RV2226	hypothetical protein	Function unknown
2502081	RV2228c	hypothetical protein	Carbohydrate transport and metabolism
2503633	RV2230c	hypothetical protein	Function unknown
2509148	RV2236c	cobalamin biosynthesis protein CobD	Coenzyme transport and metabolism
2509730	RV2237	hypothetical protein	Function unknown
2518927	RV2245	3-oxoacyl-ACP synthase	Lipid transport and metabolism
2521350	RV2247	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
2529690	RV2256c	hypothetical protein	#N/A
2531752	RV2258c	transcriptional regulator	Coenzyme transport and metabolism



2558426	RV2285	acyltransferase, WS/DGAT/MGAT	General function prediction only
2573766	IGS	intergenic space or other non-protein-coding region	#N/A
2586137	RV2314c	hypothetical protein	General function prediction only
2598410	RV2326c	cobalt/nickel transporter ATP-binding protein	General function prediction only
2612642	RV2337c	hypothetical protein	#N/A
2621068	RV2343c	DNA primase	Replication, recombination and repair
2642385	RV2360c	hypothetical protein	#N/A
2654381	RV2374c	heat-inducible transcription repressor hrcA	Transcription
2656235	RV2377c	protein mbtH	Function unknown
2660329	RV2379c	amino acid adenylation domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
2680668	RV2386A	hypothetical protein	#N/A
2695388	RV2398c	sulfate ABC transporter, permease CysW	Inorganic ion transport and metabolism
2713805	IGS	intergenic space or other non-protein-coding region	#N/A
2734084	RV2436	ribokinase	Carbohydrate transport and metabolism
2736444	RV2438c	glutamine-dependent NAD(+) synthetase	Coenzyme transport and metabolism
2745899	RV2446c	hypothetical protein	#N/A
2752708	IGS	intergenic space or other non-protein-coding region	#N/A
2754927	RV2455c	2-oxoacid:acceptor oxidoreductase subunit alpha	Energy production and conversion
2760162	RV2458	homocysteine S-methyltransferase	Amino acid transport and metabolism
2775052	RV2471	alpha-glucosidase	Carbohydrate transport and metabolism
2779146	RV2476c	glutamate dehydrogenase	Amino acid transport and metabolism
2783931	RV2477c	ChvD family ATP-binding cassette protein	General function prediction only
2786962	RV2482c	glycerol-3-phosphate acyltransferase	Lipid transport and metabolism
2805163	RV2490c	PE-PGRS family protein	#N/A
2814971	RV2501c	acetyl/propionyl-CoA carboxylase subunit alpha	Lipid transport and metabolism
2818847	RV2503c	3-oxoacid CoA-transferase subunit B	Lipid transport and metabolism
2821352	RV2505c	fatty-acyl-CoA synthase	Lipid transport and metabolism
2821622	IGS	intergenic space or other non-protein-coding region	#N/A
2827994	IGS	intergenic space or other non-protein-coding region	#N/A
2828029	IGS	intergenic space or other non-protein-coding region	#N/A
2830535	RV2513	hypothetical protein	#N/A
2855269	RV2531c	arginine decarboxylase	Amino acid transport and metabolism
2861361	RV2537c	3-dehydroquinate dehydratase	Amino acid transport and metabolism
2865770	RV2542	hypothetical protein	Replication, recombination and repair
2865892	RV2542	hypothetical protein	Replication, recombination and repair
2871058	RV2551c	leader peptidase (prepilin peptidase)/N-methyltransferase	Cell motility
2880712	RV2560	hypothetical protein	Function unknown
2881465	RV2561	hypothetical protein	#N/A
2888211	RV2566	hypothetical protein	Function unknown
2889643	RV2566	hypothetical protein	Function unknown
2891277	RV2567	hypothetical protein	Function unknown
2891738	RV2567	hypothetical protein	Function unknown
2893248	RV2568c	hypothetical protein	Function unknown
2894218	RV2569c	hypothetical protein	Amino acid transport and metabolism

2910471	RV2584c	adenine phosphoribosyltransferase	Nucleotide transport and metabolism
2911303	RV2585c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
2912304	RV2585c	peptide/nickel transporter substrate-binding protein	Amino acid transport and metabolism
2923401	RV2592c	Holliday junction DNA helicase subunit RuvB	Replication, recombination and repair
2927949	IGS	intergenic space or other non-protein-coding region	#N/A
2939383	RV2611c	phosphatidylinositol mannoside acyltransferase	Cell wall/membrane/envelope biogenesis
2939667	RV2611c	phosphatidylinositol mannoside acyltransferase	Cell wall/membrane/envelope biogenesis
2945177	IGS	intergenic space or other non-protein-coding region	#N/A
2960241	RV2634c	PE-PGRS family protein	#N/A
2966455	IGS	intergenic space or other non-protein-coding region	#N/A
2968923	IGS	intergenic space or other non-protein-coding region	#N/A
2974943	RV2650c	hypothetical protein	General function prediction only
2975293	RV2651c	HK97 family phage prohead protease	General function prediction only
2980921	IGS	intergenic space or other non-protein-coding region	#N/A
2984750	RV2668	hypothetical protein	#N/A
2987651	IGS	intergenic space or other non-protein-coding region	#N/A
3004101	RV2687c	fluoroquinolones export permease	#N/A
3005195	RV2688c	fluoroquinolones export ATP-binding protein	Defense mechanisms
3009702	RV2691	trk system potassium uptake protein TrkA	Inorganic ion transport and metabolism
3017475	RV2702	polyphosphate glucokinase	Transcription
3027592	RV2714	hypothetical protein	#N/A
3029620	RV2716	phenazine biosynthesis protein PhzF family protein	General function prediction only
3031682	IGS	intergenic space or other non-protein-coding region	#N/A
3033436	RV2721c	hypothetical protein	Cell wall/membrane/envelope biogenesis
3041881	RV2729c	hypothetical protein	#N/A
3054091	RV2741	PE-PGRS family protein	#N/A
3054331	RV2741	PE-PGRS family protein	#N/A
3061625	RV2748c	DNA translocase ftsK	Cell cycle control, cell division, chromosome partitioning
3073878	RV2764c	thymidylate synthase	Nucleotide transport and metabolism
3077049	RV2768c	PPE family protein	Cell motility
3080805	RV2771c	hypothetical protein	General function prediction only
3086798	IGS	intergenic space or other non-protein-coding region	#N/A
3094861	RV2786c	riboflavin biosynthesis protein RibF	Coenzyme transport and metabolism
3103692	RV2794c	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
3108065	RV2799	hypothetical protein	#N/A
3108684	RV2800	hydrolase	General function prediction only
3111472	RV2802c	hypothetical protein	Function unknown
3111699	RV2802c	hypothetical protein	Function unknown
3118010	RV2812	transposase	Replication, recombination and repair
3119703	IGS	intergenic space or other non-protein-coding region	#N/A
3120189	IGS	intergenic space or other non-protein-coding region	#N/A
3122094	RV2815Ac	hypothetical protein	#N/A
3122106	RV2815Ac	hypothetical protein	#N/A
3133546	RV2825c	hypothetical protein	Function unknown

3136345	RV2828Ac	hypothetical protein	#N/A
3137068	RV2830c	antitoxin	Cell cycle control, cell division, chromosome partitioning
3147180	RV2839c	translation initiation factor IF-2	Translation, ribosomal structure and biogenesis
3162815	RV2853	PE-PGRS family protein	Function unknown
3165084	RV2854	hypothetical protein	Lipid transport and metabolism
3175712	RV2864c	penicillin-binding lipoprotein	Cell wall/membrane/envelope biogenesis
3177894	RV2866	toxin	Translation, ribosomal structure and biogenesis
3178455	IGS	intergenic space or other non-protein-coding region	#N/A
3186870	RV2874	protein dipZ	Post-translational modification, protein turnover, and chaperones
3188596	RV2877c	hypothetical protein	Post-translational modification, protein turnover, and chaperones
3191037	RV2881c	phosphatidate cytidyltransferase	Lipid transport and metabolism
3193303	IGS	intergenic space or other non-protein-coding region	#N/A
3218353	RV2911	D-alanyl-D-alanine carboxypeptidase	Cell wall/membrane/envelope biogenesis
3226191	RV2916c	signal recognition particle protein	Intracellular trafficking, secretion, and vesicular transport
3228153	RV2917	hypothetical protein	Transcription
3232825	IGS	intergenic space or other non-protein-coding region	#N/A
3236240	RV2922c	chromosome partition protein smc	Cell cycle control, cell division, chromosome partitioning
3243640	IGS	intergenic space or other non-protein-coding region	#N/A
3256504	RV2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3260311	RV2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3261390	RV2933	phenolphthiocerol synthesis type-I polyketide synthase C	Secondary metabolites biosynthesis, transport, and catabolism
3269591	RV2935	non-ribosomal peptide synthetase domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
3270794	RV2935	non-ribosomal peptide synthetase domain-containing protein	Secondary metabolites biosynthesis, transport, and catabolism
3296853	RV2947c	polyketide synthase 1/15	Secondary metabolites biosynthesis, transport, and catabolism
3302501	IGS	intergenic space or other non-protein-coding region	#N/A
3308616	IGS	intergenic space or other non-protein-coding region	#N/A
3314639	RV2962c	PGL/p-HBAD biosynthesis rhamnosyltransferase	Carbohydrate transport and metabolism
3336209	RV2980	hypothetical protein	#N/A
3336597	IGS	intergenic space or other non-protein-coding region	#N/A
3336835	RV2981c	D-alanine--D-alanine ligase	Cell wall/membrane/envelope biogenesis
3338613	RV2982c	glycerol-3-phosphate dehydrogenase [NAD(P)+]	Energy production and conversion
3351936	RV2994	hypothetical protein	Carbohydrate transport and metabolism
3354906	RV2996c	D-3-phosphoglycerate dehydrogenase	Coenzyme transport and metabolism
3358245	RV2999	lipoprotein LppY	#N/A
3363348	IGS	intergenic space or other non-protein-coding region	#N/A
3367775	RV3009c	aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	Translation, ribosomal structure and biogenesis
3402827	RV3042c	phosphoserine phosphatase	Amino acid transport and metabolism
3413590	RV3051c	ribonucleoside-diphosphate reductase subunit alpha	Nucleotide transport and metabolism
3418325	RV3057c	short-chain type alcohol dehydrogenase/reductase	General function prediction only

3418327	RV3057c	short-chain type alcohol dehydrogenase/reductase	General function prediction only
3423181	RV3060c	GntR family transcriptional regulator	Transcription
3425136	RV3061c	acyl-CoA dehydrogenase fadE22	Lipid transport and metabolism
3425139	RV3061c	acyl-CoA dehydrogenase fadE22	Lipid transport and metabolism
3425851	RV3062	DNA ligase I	Replication, recombination and repair
3426792	RV3062	DNA ligase I	Replication, recombination and repair
3428914	RV3063	carbon starvation protein	Signal transduction mechanisms
3429199	RV3063	carbon starvation protein	Signal transduction mechanisms
3440461	RV3077	hydrolase	Inorganic ion transport and metabolism
3440465	RV3077	hydrolase	Inorganic ion transport and metabolism
3454983	RV3088	acyltransferase, WS/DGAT/MGAT	#N/A
3456292	RV3089	o-succinylbenzoate--CoA ligase	Lipid transport and metabolism
3456663	RV3089	o-succinylbenzoate--CoA ligase	Lipid transport and metabolism
3460983	RV3092c	hypothetical protein	Function unknown
3462132	RV3093c	oxidoreductase	Energy production and conversion
3486974	RV3121	cytochrome P450 141 cyp141	Secondary metabolites biosynthesis, transport, and catabolism
3489848	RV3124	hypothetical protein	Signal transduction mechanisms
3491390	RV3125c	PPE family protein	Cell motility
3503892	RV3137	histidinol-phosphatase	Carbohydrate transport and metabolism
3518164	RV3151	NADH-quinone oxidoreductase subunit G	Energy production and conversion
3518552	RV3151	NADH-quinone oxidoreductase subunit G	Energy production and conversion
3537745	RV3169	hypothetical protein	#N/A
3542259	RV3174	oxidoreductase	General function prediction only
3548638	RV3179	ATPase AAA+ superfamily protein	General function prediction only
3556272	RV3190c	hypothetical protein	#N/A
3564397	RV3195	hypothetical protein	Function unknown
3581410	RV3204	methyltransferase	Replication, recombination and repair
3591060	RV3213c	chromosome partitioning protein	Cell cycle control, cell division, chromosome partitioning
3594121	RV3217c	hypothetical protein	#N/A
3604818	RV3228	ribosome small subunit-dependent GTPase A	General function prediction only
3607020	RV3229c	stearoyl-CoA 9-desaturase	Lipid transport and metabolism
3607098	IGS	intergenic space or other non-protein-coding region	#N/A
3610438	RV3234c	acyltransferase, WS/DGAT/MGAT	#N/A
3614979	RV3239c	drug:H <sup>+</sup> antiporter-2 (14 Spanner) (DHA2) family drug resistance MFS transporter	General function prediction only
3622438	RV3243c	hypothetical protein	#N/A
3625062	RV3245c	sensor histidine kinase mtrB	Signal transduction mechanisms
3643389	RV3263	methyltransferase	Replication, recombination and repair
3643627	RV3263	methyltransferase	Replication, recombination and repair
3682155	RV3298c	polyhydroxybutyrate depolymerase	Secondary metabolites biosynthesis, transport, and catabolism
3684086	RV3299c	arylsulfatase	Inorganic ion transport and metabolism
3686203	RV3300c	hypothetical protein	Translation, ribosomal structure and biogenesis
3689520	RV3303c	NAD(P)H dehydrogenase (quinone)	Energy production and conversion

3704767	RV3317	succinate dehydrogenase hydrophobic membrane anchor protein	Energy production and conversion
3710553	RV3324A	4a-hydroxytetrahydrobiopterin dehydratase	#N/A
3714382	RV3328c	sigma-70 family RNA polymerase sigma factor	Transcription
3714928	RV3329	hypothetical protein	#N/A
3721977	RV3335c	hypothetical protein	Function unknown
3726058	RV3339Ac	hypothetical protein	#N/A
3736799	RV3343c	PPE family protein	Cell motility
3740084	RV3345c	hypothetical protein	#N/A
3740664	RV3345c	hypothetical protein	#N/A
3740676	RV3345c	hypothetical protein	#N/A
3740684	RV3345c	hypothetical protein	#N/A
3746580	RV3347c	hypothetical protein	Cell motility
3776877	RV3365c	hypothetical protein	Signal transduction mechanisms
3778392	IGS	intergenic space or other non-protein-coding region	#N/A
3795064	IGS	intergenic space or other non-protein-coding region	#N/A
3798275	RV3383c	geranylgeranyl diphosphate synthase, type I	Coenzyme transport and metabolism
3811507	RV3395c	hypothetical protein	#N/A
3813475	RV3396c	GMP synthase	Nucleotide transport and metabolism
3817297	RV3399	hypothetical protein	Secondary metabolites biosynthesis, transport, and catabolism
3823339	RV3403c	hypothetical protein	Function unknown
3825740	RV3406	taurine dioxygenase	Secondary metabolites biosynthesis, transport, and catabolism
3826864	RV3408	toxin	General function prediction only
3829950	RV3410c	IMP dehydrogenase	Nucleotide transport and metabolism
3839051	RV3420c	ribosomal-protein-alanine acetyltransferase	General function prediction only
3847864	RV3429/RV3430 c	PPE family protein//RNA-directed DNA polymerase	#N/A
3850377	IGS	intergenic space or other non-protein-coding region	#N/A
3859556	RV3439c	hypothetical protein	RNA processing and modification
3860073	RV3440c	hypothetical protein	#N/A
3865174	RV3447c	ESX-4 secretion system protein eccC4	Cell cycle control, cell division, chromosome partitioning
3869250	RV3448	ESX-4 secretion system protein eccD4	#N/A
3873571	RV3451	cutinase	#N/A
3877600	RV3456c	50S ribosomal protein L17	Translation, ribosomal structure and biogenesis
3885085	RV3467	hypothetical protein	#N/A
3886065	RV3468c	dTDP-glucose 4,6-dehydratase	Cell wall/membrane/envelope biogenesis
3892771	RV3476c	MFS transporter, MHS family, alpha-ketoglutarate permease	Carbohydrate transport and metabolism
3892850	RV3476c	MFS transporter, MHS family, alpha-ketoglutarate permease	Carbohydrate transport and metabolism
3894211	IGS	intergenic space or other non-protein-coding region	#N/A
3895870	IGS	intergenic space or other non-protein-coding region	#N/A
3898587	RV3479	hypothetical protein	General function prediction only
3903709	RV3484	hypothetical protein	Transcription
3908266	RV3489	hypothetical protein	#N/A
3928426	RV3507	PE-PGRS family protein	Function unknown

3931075	IGS	intergenic space or other non-protein-coding region	#N/A
3936940	IGS	intergenic space or other non-protein-coding region	#N/A
3941080	RV3511	PE-PGRS family protein	Function unknown
3942660	IGS	intergenic space or other non-protein-coding region	#N/A
3942819	IGS	intergenic space or other non-protein-coding region	#N/A
3943198	IGS	intergenic space or other non-protein-coding region	#N/A
3952979	RV3516	enoyl-CoA hydratase	Lipid transport and metabolism
3958582	RV3521	hypothetical protein	General function prediction only
3959597	RV3522	hypothetical protein	Lipid transport and metabolism
3962617	IGS	intergenic space or other non-protein-coding region	#N/A
3974133	RV3535c	acetaldehyde dehydrogenase	Secondary metabolites biosynthesis, transport, and catabolism
3983450	RV3544c	hypothetical protein	Lipid transport and metabolism
4005786	RV3564	hypothetical protein	Lipid transport and metabolism
4024452	RV3581c	2-C-methyl-D-erythritol 2,4-cyclo diphosphate synthase	Lipid transport and metabolism
4027078	RV3585	DNA repair protein RadA	Post-translational modification, protein turnover, and chaperones
4030722	RV3589	A/G-specific adenine glycosylase	Replication, recombination and repair
4035006	RV3593	hypothetical protein	Defense mechanisms
4037462	RV3595c	PE-PGRS family protein	#N/A
4038466	RV3596c	ATP-dependent Clp protease ATP-binding subunit ClpC	Post-translational modification, protein turnover, and chaperones
4053229	RV3611	hypothetical protein	#N/A
4055980	RV3616c	ESX-1 secretion-associated protein EspA	#N/A
4060083	IGS	intergenic space or other non-protein-coding region	#N/A
4069471	RV3630	hypothetical protein	General function prediction only
4076136	IGS	intergenic space or other non-protein-coding region	#N/A
4088525	RV3648c	CspA family cold shock protein	Transcription
4099459	IGS	intergenic space or other non-protein-coding region	#N/A
4101153	IGS	intergenic space or other non-protein-coding region	#N/A
4111481	RV3669	hypothetical protein	#N/A
4131322	RV3689	hypothetical protein	RNA processing and modification
4135290	RV3693	hypothetical protein	General function prediction only
4136759	RV3694c	hypothetical protein	Function unknown
4150826	RV3707c	hypothetical protein	#N/A
4156277	RV3711c	exonuclease, DNA polymerase III, epsilon subunit	Replication, recombination and repair
4159675	RV3714c	hypothetical protein	#N/A
4162517	RV3719	hypothetical protein	Energy production and conversion
4164122	RV3720	cyclopropane-fatty-acyl-phospholipid synthase	#N/A
4174742	RV3727	oxidoreductase	Function unknown
4182873	RV3731	DNA ligase (ATP)	Replication, recombination and repair
4187663	RV3736	hypothetical protein	Transcription
4187995	RV3737	hypothetical protein	Function unknown
4188834	RV3737	hypothetical protein	Function unknown
4204619	RV3759c	osmoprotectant transporter substrate-binding protein	Cell wall/membrane/envelope biogenesis
4210452	RV3764c	two-component system, OmpR family, sensor kinase	Signal transduction mechanisms

4220352	RV3775	hypothetical protein	Defense mechanisms
4222251	IGS	intergenic space or other non-protein-coding region	#N/A
4223060	RV3777	oxidoreductase	Energy production and conversion
4225144	IGS	intergenic space or other non-protein-coding region	#N/A
4233477	RV3786c	hypothetical protein	Cell wall/membrane/envelope biogenesis
4241213	RV3793	arabinosyltransferase C	#N/A
4242360	RV3793	arabinosyltransferase C	#N/A
4242821	RV3793	arabinosyltransferase C	#N/A
4247647	RV3795	arabinosyltransferase B	#N/A
4256100	RV3799c	propionyl-CoA carboxylase subunit beta	Lipid transport and metabolism
4257398	RV3800c	polyketide synthase 13	Secondary metabolites biosynthesis, transport, and catabolism
4264396	RV3802c	hypothetical protein	#N/A
4287189	RV3822	hypothetical protein	#N/A
4291596	RV3823c	membrane transporter mmpL8	General function prediction only
4293250	RV3824c	SL659 acyltransferase papA1	Secondary metabolites biosynthesis, transport, and catabolism
4299658	RV3825c	phthioceranic/hydroxyphthioceranic acid synthase	Secondary metabolites biosynthesis, transport, and catabolism
4302214	RV3827c	transposase	Replication, recombination and repair
4306333	RV3831	hypothetical protein	#N/A
4307357	RV3833	AraC family transcriptional regulator	Transcription
4315562	RV3842c	glycerophosphoryl diester phosphodiesterase	Energy production and conversion
4316224	RV3843c	hypothetical protein	#N/A
4320228	IGS	intergenic space or other non-protein-coding region	#N/A
4338909	IGS	intergenic space or other non-protein-coding region	#N/A
4340507	RV3864	hypothetical protein	#N/A
4341149	RV3864	hypothetical protein	#N/A
4351216	RV3872	PE family protein	#N/A
4356287	RV3877	ESX-1 secretion system protein eccD1	#N/A
4363246	RV3882c	ESX-1 secretion system protein EccE1	#N/A
4366449	RV3884c	ESX-2 secretion system protein EccA2	Post-translational modification, protein turnover, and chaperones
4373673	RV3889c	ESX-2 secretion-associated protein EspG2	#N/A
4375805	RV3892c	PPE family protein	#N/A
4376275	IGS	intergenic space or other non-protein-coding region	#N/A
4379857	RV3894c	type VII secretion protein EccCa	Cell cycle control, cell division, chromosome partitioning
4382151	RV3896c	hypothetical protein	General function prediction only
4382231	RV3896c	hypothetical protein	General function prediction only
4382452	RV3896c	hypothetical protein	General function prediction only
4385707	RV3900c	hypothetical protein	#N/A
4386789	RV3901c	hypothetical protein	#N/A
4390930	RV3905c	WXG100 family type VII secretion target	Function unknown
4391730	RV3906c	hypothetical protein	#N/A
4400035	RV3910	integral membrane protein MviN	General function prediction only
4405133	RV3916c	hypothetical protein	#N/A
4408128	RV3919c	ribosomal RNA small subunit methyltransferase G	#N/A

4408332	RV3919c	ribosomal RNA small subunit methyltransferase G	#N/A
---------	---------	---	------